An empirical and computational investigation of variable outcomes in Autism Spectrum Disorder

An Empirical and Computational Investigation of Variable Outcomes in Autism Spectrum Disorder

Rachael Davis

Doctor of Psychology

Birkbeck, University of London

2016

Declaration

I, Rachael Davis, confirm that all the work presented in this thesis is my own.

The first part of my PhD project is based on data from the Phase 1 and Phase 2 cohorts of the British Autism Study of Infant Siblings (BASIS). This involved secondary data analyses on datasets that were already available. The second part of the thesis utilised the method of computational modelling. Three populations were analysed here. The first (known as the high-risk population) originated from the Over-Pruning account of ASD (Thomas, Knowland & Karmiloff-Smith, 2011). However, I also created two further computational populations specifically for this thesis, which I devised.

Based on the work presented in this thesis, I will be starting a 6-month award from the Birkbeck Wellcome Trust Institutional Strategic Support Fund to publish articles based on the following chapters.

4th February 2016

Acknowledgements

The work presented in this thesis has been supported by a Bloomsbury Colleges Scholarship to R. Davis, the BASIS funding consortium led by Autistica (www.basisnetwork.org), Autism Speaks (PI M. H. Johnson, Grant number 1292) and a UK Medical Research Council Programme Grant.

I would like to thank my two supervisors, Professor Michael Thomas and Professor Tony Charman for their guidance, encouragement and reassurance. Without their support and inspiration over the last three years I would not have completed this PhD. I am grateful to the members of the Centre for Brain and Cognitive Development, Birkbeck College who have helped me throughout my PhD, and the Developmental Neurocognition Lab for their input, support and discussions over the last three years. I would also like to thank Dr Greg Pasco for his assistance in utilising the BASIS datasets, and Kier Finnegan for his encouragement and statistical guidance.

I am especially grateful to my parents, Angela and Robert Davis, who have patiently reassured me and encouraged me throughout. I would also like to thank James Swatton, who has continually motivated and supported me throughout the last three years, and particularly throughout the write up of my thesis.

Abstract

This thesis had two aims. The first was to investigate variability observed in the profiles of young children with autism spectrum disorder (ASD), and our ability to predict this variability based on measures in infancy. The second aim was to identify the underlying mechanisms that generate this variability. I combined analyses of clinical data sets and data from computational models to investigate the influences shaping atypical developmental trajectories in ASD.

The first aim was addressed using secondary data analysis from a prospective longitudinal dataset, the British Autism Study of Infant Siblings. Clinical, behavioural, and parental report data were collected on 54 infants at risk of ASD (by virtue of having an older sibling with the disorder) and 50 low-risk controls at 7, 14, 24 and 36 months. Chapter 2 investigates whether variability differed at a group level, evaluating whether heterogeneity was exaggerated in high-risk groups versus low-risk controls. Cognitive variability scores distinguished infants with ASD at 36 months. Intra-subject variability was then assessed. A more uneven cognitive profile at 24 months was predictive of lower cognitive abilities at 36 months in high-risk infants overall. In Chapter 3, behavioural measures at 14 months were identified as predictors of diagnostic outcome at 36 months in high-risk infants. Initial results highlighted the importance of environmental factors and social and communicative performance. The predictive power of the subsequent statistical regression equations was validated against recently available data from Phase 2 of the BASIS study, with 125 at-risk infants, demonstrating 71% specificity and 81% sensitivity in predicting ASD characteristics at 24 months.

In the second half of the thesis, potential mechanisms generating variability in ASD behavioural profiles were investigated via computational modelling. Thomas, Knowland and Karmiloff-

Smith (2011) developed a computational model targeting the regressive sub-type of autism based on the hypothesis that regression could be caused by Over-Pruning of brain connectivity. In Chapter 4, this model is extended to capture other observed developmental trajectories in ASD. Regressive and non-regressive subgroups were identified, and each was reliably distinguished by a distinct pattern of neurocomputational parameters. Regression and early onset of pruning were indicative of poorer developmental outcomes overall. Non-regressive subgroups, both typical and atypical, were then used to investigate response to remediation via behavioural intervention. The simulation work represents the first application of population-level models of atypical development to intervention. Small but reliable intervention effects were identified, following a discrete phase of intervention. However, the results indicated a limited scope to intervene, with the greater success using compensatory rather than normalisation techniques.

The overall results are discussed with reference to the need for convergent methods to shed light on the constraints shaping atypical developmental trajectories in ASD.

Contents

List of Tables

Figures

Chapter 1

1.1 Introduction

Autism Spectrum Disorder (ASD) is a pervasive developmental disorder, characterised by impairments in social and communicative domains, restrictive or repetitive behaviours and interests and atypical sensory behaviours and/or interests (DSM-5, American Psychiatric Association, 2013). The prevalence of ASD is estimated to be around 1 in 100 in the general population (Brugha et al. 2011) and, due to the high heritability of the disorder, recurrence rates are estimated to be around 10% (Constantino et al., 2010) for those who have a first-degree relative with a diagnosis. Such elevated prevalence levels combined with the fact that diagnosis is rarely received before the age of 3 years (Centres for Disease Control, 2012) has made cognitive and behavioural profiling and understanding the emergence of symptoms problematic. Therefore, the implementation of longitudinal studies, following at-risk infants at multiple time points across infancy has provided insight not only into the development, but also cognitive and behavioural changes in infants with ASD (Szatmari et al., 2016; Jones et al., 2014; Zwaigenbaum et al., 2009). Identifying possible developmental processes can enable earlier diagnosis and in turn support earlier intervention approaches (Zwaigenbaum et al., 2015).

The first aim of this thesis is to investigate variability observed in the profiles of infants at risk of developing ASD and to examine the underlying processes that influence variability in developmental trajectories. The second aim is to predict variability using behavioural measures from early infancy. Two approaches will be implemented to elucidate these research questions. First, cognitive, behavioural and environmental measures from 7- and

14-month evaluations will be analysed using data from *the British Autism Study of Infant Siblings* (BASIS) (www.basisnetwork.org), a longitudinal prospective study of infants with a familial risk of developing ASD due to having an older sibling with the disorder. Identifying the early influences on atypical developmental trajectories in ASD could benefit the detection of the earliest emerging symptoms of ASD and importantly, facilitate effective interventions at an earlier stage in infancy.

The second approach utilises populations from a neural network computational model of ASD. Computational models can be powerful tools in advancing theories of cognitive development. However, within the handful of models that attempt to identify causal mechanisms in ASD, very few are true developmental models. Only one model, the Over-Pruning hypothesis of autism (Thomas, Davis, Karmiloff-Smith, Knowland & Charman, 2015) has utilised population modelling to investigate variability in developmental trajectories and long-term outcomes. I develop and analyse novel populations from this computational model as an alternative approach to understanding the mechanisms explaining heterogeneity identified in the onset, severity and outcomes in ASD. Bringing together these elements, I use clinical data to investigate variability in infants, and empirically test a number of predictions from the Over-Pruning account.

The current chapter begins with a review of the methodologies employed in the emergence of ASD symptoms, that is retrospective versus prospective and longitudinal designs. I present a summary of the most current and relevant theoretical positions of ASD, with a focus on a neuroconstructivist account of developmental disorders. I discuss variability, and in particular whether autism is comprised of one, or multiple disorders, and present findings

from genetic and behavioural research that focus on this question. A number of methods have also been used in an effort to further understand variability and heterogeneity, including the identification of subgroups within ASD based on developmental trajectories, the search for biomarkers and use of statistical models of variability. I discuss the state of intervention research thus far, and the questions that have arisen from such studies. Finally, I consider the value of computational models in developmental research in terms of understanding mechanistic sources of variability, and analysing intervention effects in delayed or impaired networks. The Over-Pruning account of ASD is also introduced.

1.2 Retrospective vs. prospective methods

Heterogeneity and variability of clinical presentation among individuals with ASD greatly contributes to the challenge of early diagnosis. However, following the progression of ASD from the earliest phases in development is essential in order to identify the developmental mechanisms underlying the disorder, and to provide the earliest opportunities for intervention. In order to identify early behavioural and cognitive atypicalities, two general methodologies have been developed (Yirmiya & Charman, 2010). The first, known as a retrospective design, involves the assessment of evidence from past behaviours, usually through the evaluation of home videos or parental reports. The second, known as a prospective design, measures infant development across a range of cognitive and behavioural domains longitudinally, at multiple time points. Retrospective designs have provided valuable insights into the identification of early signs of ASD (Zwaigenbaum et al., 2005). However, the ability to trace differences in developmental trajectories across a range of social, communicative, motor and behavioural domains using prospective methods has extended findings from retrospective research, providing more reliable results.

Retrospective designs use early evidence from children who have already received an ASD diagnosis. One of the most commonly employed methods based around this approach is the parental report, or semi-structured interview where parents are asked to recall numerous aspects of their child's early development. A number of reports have been used in such studies. These include: The Detection of Autism by Infant Sociability Interview (DAISI) (Wimpory, 2012), which identifies whether 19 aspects of social engagement that are characteristic in typically developing children were present in the first year of life; the Autism Diagnostic Interview-Revised (ADI-R) (Rutter, LeCoulter & Lord, 2003, 2008) which covers eight areas of questioning, including early development and developmental milestones, language acquisition and interests and play behaviours; and the Childhood Autism Rating Scale (Schopler, Van Bourgondien, Wellman & Love, 2010), which rates children on a scale from 1 to 4 on criteria including verbal communication, anxiety and nervousness, imitation and relationships to people.

When parents were asked about their initial concerns regarding their child with ASD, 30% to 50% of parents recalled atypical behaviours within the first year of infancy (Zwaigenbaum et al., 2005). For example, Gillberg et al. (1990) assessed 28 children with ASD at 3 years. Parental interviews were conducted using the Autism Behaviour Checklist (ABC; Krug, Arick & Almond, 1980), and focused on the onset of early symptoms and family history. Parents reported atypicalities in hearing and a lack of play behaviours in the first 12 months of infancy. Other retrospective studies have reported similar findings in terms of play (Rogers & DiLalla, 1990; Dahlgren & Gillberg, 1989; Ohta et al., 1987), a lack of response to parental voices (Hoshino et al., 1987) and inconsistent or poor eye contact (Hoshino et al., 1987; Rogers & DiLalla, 1990).

It is not until the second year of life that parents report core symptoms such as speech delay (Rogers & DiLalla, 1990) or restricted and repetitive interests (Stone et al., 1994). This method is arguably one of the simplest to implement because short interviews are often the only requirement for participating parents. Such methods, however, have several significant disadvantages.

Firstly, parents are subject to recall bias. For example, if a child has received a diagnosis, parents can produce biased recollections that are consistent with diagnostic information that they have already received. Retrieving information about behaviours that occurred up to a decade prior to the parental reports can also be prone to distortion and misremembering, increasing the risk of reporting significant inaccuracies. By utilising clinical reports alone, researchers are limited by the explicit, selective and often positively biased behaviours identified by parents, in comparison with longitudinal clinical observations (Jones et al., 2014), which provide an opportunity to systematically measure the development of ASD.

A second retrospective method is home videotape analysis. A major advantage to this method is that it allows a child's behaviour to be assessed in an impartial setting by unbiased observers. Secondly, control groups can be used to compare children with ASD to typically developing individuals, and coding schemes are used to analyse video data, which allows for the assessment of behaviours using standardised criteria. Studies of early home videos have identified a number of atypical behaviours in children who are later diagnosed ASD in comparison with typically developing children (Adrien et al., 1992; Werner et al., 2000; Osterling & Dawson, 1994; Baranek et al., 1999). Using home video, a number of atypical behaviours have been identified in the first year of life that are associated with a later diagnosis. Children with ASD are distinguished at this time by a lack of social smiling and

facial expressions (Adrien et al., 1992), fewer joint attention behaviours (Osterling & Dawson, 1994) and atypical posture and patterns of movement (Adrien et al., 1993). Baranek et al. (1999) compared 10 children with ASD, 10 with developmental disabilities and 11 typically developing children using home videos between 9 and 12 months. A total of nine items in combination (including response to name, anticipatory posture and orienting behaviours) were found to discriminate the ASD group with an accuracy level of 93%.

However, home video data are still problematic in that children are more likely to be filmed on special occasions (e.g. first birthday parties), or when they are exhibiting positive behaviours. Thus, a child's typical daily behaviours may not be reflected. It is this selectivity that causes problems with generalisation of findings. Overall, retrospective studies have yielded important insights into the atypical behaviours that constitute some of the primary signs of ASD in infancy. However, the difficulty in detecting subtle behavioural changes has meant that the identification of predictive atypicalities and risk markers has been relatively unsuccessful. It has been suggested that retrospective studies can be used more appropriately as a guide to the development of screening procedures, and that prospective studies can validate such findings (Zwaigenbaum et al., 2007).

Prospective longitudinal studies measure development in behavioural and cognitive domains in infants at multiple time points across infancy and early childhood. Generally, prospective studies of ASD utilise high-risk samples; that is a group that are at a higher risk of developing the disorder. Most commonly, this refers to the presence of an older sibling with an ASD diagnosis. Siblings of children with ASD are at a substantially higher risk of developing the disorder than the general population (or community populations who do not have a family history of ASD). Recurrence rates are shown to range from 5% - 10% in community samples

(Constantino et al., 2010; Bolton et al., 1994), and as high as 20% in high-risk research samples (Ozonoff et al., 2011). Jones, Gliga, Bedford, Charman and Johnson (2014) suggest the fluctuation in recurrence rates could be due to a combination of failing to identify milder forms of ASD in community population samples (often these individuals are recruited via community routes such as volunteer schemes and not, for example, through clinical services), and parents choosing not to have more children if one child has already been diagnosed with ASD. In general, high-risk prospective studies measure the development of younger siblings of children who have already received a diagnosis of ASD. These children are tracked across the first three years of life and assessed at multiple time points. At the final time point (commonly either 24 or 36 months) a diagnostic outcome is decided for each child. That is, children either receive or do not receive a diagnosis of ASD. Behavioural, environmental and cognitive comparisons can be made between high-risk infants with ASD, low-risk infants who have no familial risk of ASD, and with high-risk infants who were not diagnosed with ASD. Comparing the developmental trajectories of children in the high-risk typically developing groups, high-risk ASD groups and low-risk groups allows for the identification of early atypical behavioural and cognitive markers that could predict developmental outcome at 3 years, as well as potential environmental risk and protective factors.

In addition to the high-risk population who develop ASD, a second population known as the broader autism phenotype (BAP) can be studied. This group exhibit sub-clinical symptoms of ASD and are more likely to be identified in a high-risk family. The BAP was first reported by Bolton (1994) who identified a higher number of social and communicative impairments in individuals with first-degree relatives with ASD when compared with individuals with an equivalent relative with Down syndrome. Consistent with these findings, Pickles et al. (2000) and Piven et al. (1997) identified higher rates of personality characteristics associated

with ASD (such as rigidity, and speech and pragmatic language deficits) in parents who had two children with ASD compared with parents that had children with Down syndrome probands. Furthermore, Ozonoff et al. (2014) investigated the prevalence of the BAP in a high-risk sample of 294 infant siblings in order to identify the frequency and the age at which the BAP was first evident. A total of 28% of the high-risk group were classified as non-typically developing at 26 months of age. Moreover, whilst non-typically developing infants were indistinguishable from low-risk typically developing infants at 6 months, significant differences were identified between the groups in language, motor, social and cognitive domains by 12 months. Thus far, prospective research has enabled the identification of behavioural markers within the second year of infancy across social, communicative and sensory domains. Furthermore, the identification of subgroups marked by atypicality in cognitive and developmental domains has begun to address the problem of heterogeneity both within and between individuals with ASD (see Chapter 3 for a detailed review of prospective high-risk findings thus far).

1.3 Theories of ASD

A vast number of causal theories of ASD have been proposed, particularly over the last 30 years (Ronald & Happé, 2008; Lai, Lombardo & Baron-Cohen, 2014). To provide a comprehensive overview of the entire field and its theories would be too great a scope for this thesis; therefore, the most relevant theoretical frameworks and theories for this thesis will be discussed.

Many theories propose that primary deficits in ASD are either social or cognitive in origin. Furthermore, the vast majority of theories of ASD propose a single causal factor that could

cause cascading effects across a wide range of domains later in development. These theories include the Theory of Mind (Baron-Cohen, Leslie & Frith, 1985; Baron Cohen, 1985), the Social Brain theory (Hadjikhani, Joseph, Snyder & Tager-Flusberg, 2006), atypicalities in social orienting (e.g. Dawson, Melzoff, Osterling & Rinaldi, 1998; Halit, Csibra, Volein & Johnson, 2004) or social processing (e.g. Mundy & Neal, 2001). Alternatively, there have been proposals of impairments in global processing (known as the Weak Central Coherence theory (Happé & Frith, 1994; Frith, 2003), an account of Enhanced Perceptual Functioning (Mottron & Burack, 2001; Mottron, Burack, Iarocci & Enns, 2003) and theories of over-connectivity (Rubenstein & Merzenich, 2003) and under-connectivity (Keary et al., 2009; Courchesne et al., 2001; Casanova et al., 2006).

It has been argued that many theories fail to consider the developmental process itself when hypothesising about the causes of ASD (Karmiloff-Smith, 1998). Many hypotheses, such as the Theory of Mind account, view the brain as a static entity and argue for innate modularity. It is their view that impairments are the result of atypicalities within particular domain-specific cognitive modules, and frequently, behavioural data is characterised in terms of intact and spared abilities (Karmiloff-Smith, 1998, 2009). Developmental models, on the other hand, stress the significance of understanding the role of developmental processes, plasticity, and identifying developmental trajectories. The emphasis lies in the developmental processes that lead to the manifestation of behaviours presented in individuals with ASD as opposed to static models, which construct theories around individual impairments. The neuroconstructivist framework (Westermann et al., 2007) for example, suggests that intelligence is the product of multidirectional interactions at the genetic, behavioural, cognitive and environmental levels. Johnson (2001) argues against the implication of genetic determination suggested in many non-developmental theories, and

instead proposes that the brain is self-structuring and changes over developmental time as a result of multiple levels of interaction.

The social motivation hypothesis (Mundy & Neal, 2001; Dawson, Webb & McPartland, 2005; Chevallier, Kohls, Troiani, Brodkin & Shultz, 2012) is an example that places developmental processes at the core of explanations for ASD. Chevallier and colleagues suggest that reduced attention towards social information may underlie many of the deficits in ASD. Specifically, they argue that ASD can be seen as an extreme case of diminished social motivation. They hypothesise that early-onset impairments in social attention initiate atypical developmental processes that ultimately result in inadequate social environments for important learning experiences. As a result, children with ASD exhibit an imbalance in attending to social and non-social stimuli, which leads to further atypicalities in social cognition. Furthermore, Jones, Webb, Estes, and Dawson (2013) suggest that the emergence of core symptoms may relate to atypical development of the neural circuitry that is involved in forming representations related to the reward value of social stimuli. Therefore, early core symptoms of ASD may characterise the failure of social-based neural systems to develop. Webb, Jones, Kelly and Dawson (2014) argue that early interventions promoting social engagement through the practise of dyadic interactions, reciprocity and communication skills could provide the necessary additional stimulation to promote the typical development of social-based neural circuitry, which in turn, could redirect brain development towards a more typical trajectory, and potentially reduce the manifestation of symptoms.

Similarly, the Interactive Specialisation account (Johnson et al., 2005) is a theory of brain development that can account for deficits across multiple domains in ASD. The account proposes that development is a product of a complex and constant interaction between

genetics, environment and neural mechanisms, and even prenatal experiences can affect the developing specialisation of neural pathways. Simply put, it is suggested that in ASD, infants may process and attend to atypical features in their environment. Over time, atypical processing will affect the experiences of the infant, causing downstream consequences in development, which would accordingly consolidate an atypical developmental trajectory. One example is that early atypical gaze behaviours could lead to less frequent eye contact and dyadic behaviour with a caregiver. This in turn could contribute to poorer communication skills and difficulties with interactions. Consequently, the infant displays an atypical developmental trajectory, as atypical connections have been strengthened by very early atypical experiences. Johnson argues that by engaging the brain at the height of neural plasticity, it may be possible to improve deficits in neural circuitry, and shift towards a more typical developmental trajectory.

1.4 Neural correlates of ASD

One of the greatest challenges in ASD research will be to identify the developmental trajectories of atypical neural mechanisms from early infancy. However, researchers suggest that the implications for understanding neural mechanisms could be extremely significant in long-term outcomes in ASD. For example, Vaccario et al. (2009) propose that an altered trajectory of brain development may be one of the most reliable biomarkers in ASD. One hope is that if neural correlates or biomarkers can be identified, it provides the potential to identify more successful screening procedures to trigger behavioural or pharmacological interventions. Here I focus on connectivity-based theories of ASD, as they are most relevant to the theories presented in this thesis, particularly the Over-Pruning account that is described below.

Connectivity theories of autism posit that individuals with ASD exhibit altered neural connectivity. However, thus far, much of the research appears to be contradictory. Researchers have assessed the roles of both long-range and short-range connectivity, in addition to identifying atypicalities in specific or widespread areas of neural functioning. In the following section I discuss theories proposing that over-connectivity (hyper-connectivity), under-connectivity (hypo-connectivity), or indeed, a combination of both are present in ASD.

In recent years, the study of connectivity in ASD has attracted widespread attention (Happé & Frith, 2002; Brock, Brown, Boucher & Rippon, 2002). However recent attempts to focus on atypical neural connectivity have proposed a deficit of hypo-connectivity. One particularly influential account, known as the 'under-connectivity theory' (Just et al., 2007, 2012) attributes ASD to a disruption of neural circuitry, leading to a reduction in anatomical and functional connectivity (where functional connectivity refers to the synchronisation of activation between multiple brain regions) in precise brain areas. Specifically, it is suggested that under-connectivity compromises the ability of the frontal cortex to communicate with temporal brain regions, therefore affecting a wide variety of behavioural performances where the frontal cortex is utilised. Using functional magnetic resonance imaging (fMRI) at a whole-brain level, functional connectivity was found to be significantly lower in individuals with ASD in the frontal and temporal cortices during a language comprehension task. Furthermore, under-synchronisation of activation during task performances has been identified in a number of tasks (Kana et al., 2006, 2009; Koshino et al., 2008; Mason et al., 2008; Tyszka et al., 2014).

Alternatively, others have argued that the neural basis of autism assumes global, atypical brain connectivity; specifically, generalised under-connectivity that can be identified across the whole brain (Brock et al., 2002; Belmonte et al., 2004; Bourgeron et al., 2009). The authors propose that atypical connectivity and synapse formation could explain both the manifestation of behavioural symptoms and neuroanatomical differences identified in individuals with ASD. In particular, long-range brain connectivity may be disrupted in ASD. They hypothesise that the enlargement of brain size in early stages of development (Langen et al., 2009; Hardan et al., 2009; Courchesne et al., 2003) increases the distance between central processing areas in the cortex, and therefore compromises interregional communication. However, very few studies have provided evidence that is consistent with the concept of global long-range hypo-connectivity. For example, whilst Monk et al. (2006) identified widespread functional under-connectivity within the default mode network (composed of several interacting brain regions associated with moral reasoning, theory of mind and episodic memory), this was accompanied by an increase in connectivity in the right temporal lobe and hippocampus. Furthermore, it has been proposed that atypical brain growth can lead to different optimal patterns of connectivity. For example, Braitenberg (2001) suggest that larger brains have a higher proportion of short-distance to long-distance connectivity, which contradicts the idea of abnormal connectivity as suggested by the hypo-connectivity account.

In contrast, other theories have demonstrated over-or hyper-connectivity in individuals with ASD. Uddin et al. (2013) identified hyper-connectivity in a number of large-scale brain networks in children (aged 7-12) with ASD when compared with age-matched controls. Using fMRI, the authors identified hyper-connectivity in salience (networks that determine which stimuli should be attended to in an environment), frontotemporal, motor and visual

networks. Moreover, when maps of each individual's salience networks were compared, children with ASD could be correctly discriminated from typically developing controls with a 78% accuracy level.

A number of fMRI studies have presented evidence of short-range or local over-connectivity in individuals with ASD. Schmitt et al. (2006) compared brain activation in 10 individuals with ASD with a control group using a motor inhibition task. The authors reported an increase in activation in the left inferior and orbital frontal gyrus in individuals with ASD. Furthermore, they reported a correlation between grey matter density and functional activation in the same area (see also Rubenstein & Merzenich, 2003; Schmitt et al., 2006; Belmonte & Yurgelen-Todd, 2003 for similar findings across different domains.)

Emerging findings however, present a more a more complex account of connectivity, with under-and over-connectivity identified in individuals with ASD across distinct areas of the brain. A new database, the Autism Brain Imaging Data Exchange (ABIDE) combines resting state functional magnetic resonance imaging (R-fMRI) scans from 17 laboratories, providing scientists with over 500 brain images from individuals with ASD between the ages of 7 and 64. The first study to utilise these data addressed the inconsistent findings surrounding connectivity in ASD. Martino et al. (2013) compared structural and phenotypic information from 360 individuals with ASD and 403 age-matched controls, focusing on the differences between whole-brain functional connectivity and a range of measures of functional brain architecture that are implicated in ASD. The authors propose that their findings resolve the debate as the type of connectivity that features in autism, identifying both hypo-connectivity and hyper-connectivity in different areas of the brain. Specifically, hypo-connectivity was identified in cortical regions, whereas hyper-connectivity was more dominant throughout the

brain, and particularly identifiable between cortical and subcortical brain regions and in interhemispheric functional connectivity. Other research combining data from the ABIDE project with functional imaging support the idea of the combination of hypo- and hyper-connectivity (Glerean et al., 2015; Nomi & Uddin, 2015).

It has recently been argued that neuroimaging studies must place findings within a developmental framework in order to reconcile the conflicting connectivity patterns seen in structural and functional research (Uddin, Superkar & Menon, 2013). In a review of functional connectivity in younger children, adolescents and adults with ASD, the authors propose that reduced functional connectivity is evident in adolescents and adults, whereas increased connectivity appears to be a feature of younger children with ASD. A second developmentally driven study of connectivity examined white matter fibre tract organisation using diffusion tensor imaging in 92 high-risk infants (Wolff et al., 2012) at 6 months, 12 months and 24 months of age. White matter tracts connect the cortex with other areas of the central nervous system, different areas in the same hemisphere, or the same cortical areas in opposite hemispheres (Luijkx & Rabou et al. 2013). The majority of fibre tracts were significantly different in high-risk infants who were later diagnosed with ASD compared with those who were typically developing. Interestingly, the pattern of tract atypicalities differed across development, with fibre tracts characterised by higher fractional anisotropy (a measure of microstructural integrity) at 6 months of age in infants with ASD, but a lower value by 24 months of age. Here, white matter atypicalities were identified at 6 months and Wolff and colleagues propose that white matter pathway changes may precede the manifestation of ASD symptoms.

Based on the variability of findings reported thus far, it may be too early to draw firm conclusions from these studies. The developmental profile of hypo- and hyper-connectivity is yet to be established, and although neural differences are reported in children with ASD, these differences are small. Moreover, many of these studies have not been replicated and verified. However, findings do highlight the necessity to move from utilising static methodologies to measuring connectivity across development. Such a move could lead to a more comprehensive understanding of the issues surrounding connectivity in ASD.

1.5 Intervention in ASD

Over the past twenty years, there has been a huge interest in developing novel and effective interventions for infants and young children with ASD. This increase in intervention trials can be related to the utilisation of infant sibling studies. A number of interventions have focused on infant siblings in order to initiate interventions at the earliest time point. In the past decade, progress has been made in the field of intervention due to the implementation of randomised control trials in intervention studies. This method is the most rigorous way of determining cause and effect relationships between outcome and treatment (Zwaigenbaum et al., 2013). Thus, the focus in this section is interventions that utilised a randomised control trial design.

One of the first randomised control interventions in ASD focused on improving socio-emotional behaviours and academic skills by utilising applied behaviour analysis (ABA) therapy (Smith, Groen & Wynn, 2000). ABA focuses on the concepts that explain how learning takes place, with the aim of increasing beneficial behaviours and decreasing behaviours that are seen to negatively affect learning. A number of learning strategies are

implemented in ABA therapy, which are customised for the needs of each individual based on their impairments. These include positive reinforcement of positive behaviours, story-based activities and using naturalistic teaching methods to improve functional skills. Intensive treatment groups received 30 hours of training for one year, which gradually reduced in hours over a further period of two years. These infants were compared with a parent-training group, who received 3 to 9 months of parent based training. The authors report that children in the ABA therapy group outperformed the parent-training group on IQ, language development and visuo-spatial tasks. Furthermore, children in the ABA treatment group were in general more likely to be assigned to less restrictive classroom placements, and were more likely to attend mainstream schools.

Many intervention programs have focused on social and communication skills that develop over the first 2 years of infancy. More recently, treatment efforts have focused on joint attention, which refers to the sharing of attention with others through the use of gestures or looking between people and objects (Kasari, et al., 2001), and symbolic play, the ability to use one object or action to represent another. Kasari, Freeman and Paparella (2006) recruited 58 children aged between 3 and 4 years who were randomised to a joint attention intervention, a symbolic play intervention or a control group. Each intervention type utilised ABA techniques such as verbal prompts and positive reinforcement, and children worked with trained experts for 30 minutes a day for 6 weeks. In each session, a targeted skill was identified for each child, which was taught using systematic prompting, reinforcement of events and corrective feedback in a semi-structured play session. After 6 weeks, both intervention groups exhibited significant improvements over the control group on certain social behaviours. Children in the joint attention group demonstrated greater levels of responsiveness in joint attention tasks, whilst the children in the symbolic play intervention

group displayed more diverse types of symbolic play in interactions with their caregiver. The authors argue that the intervention demonstrated promising results in social and communicative domains, but would need to examine long-term effects in order to establish the validity of the intervention type.

Similarly, Kasari, Gulsrud, Wong, Kwon and Locke (2008) utilised a caregiver-mediated intervention in the domain of joint attention. Intervention focused on play routines in which the caregiver could identify their child's interests and expand up on play activities. A total of 19 children from the intervention group were compared with 19 control children (where infants were assigned to a waitlist group and were given no intervention). The intervention group demonstrated significant improvements in terms of responsiveness to joint attention, exhibiting greater variation in the type of play behaviours displayed. Furthermore, acquired skills in joint attention were maintained and improvement in language abilities were found at a follow-up one-year after intervention had ceased.

Kaale, Smith and Sponheim (2012) reported analogous findings using a joint attention based intervention in a pre-school setting. The authors argue for the importance of implementing school-based interventions due to restrictions in the time available for parents to be involved in the intervention and because of the benefit of using available pre-school time. Sixty-one children aged between 29 and 60 months with a diagnosis of ASD were randomised to either 8 weeks of joint attention intervention in addition to their daily preschool program, or to their preschool program only. Intervention procedures were adapted from those used by Kasari et al. (2006) and conducted by teachers who were supervised weekly by trained clinicians. Infants in the treatment group demonstrated significantly higher levels of joint attention and joint engagement behaviours post-intervention when compared with the non-treatment group.

The effects also generalised to increased levels of joint engagement with caregivers. The authors suggest that a combination of parent-mediated and school-based joint attention intervention programs may generate the most successful long-term outcomes in joint attention and engagement. However, long-term outcomes are necessary to address whether these intervention effects will elicit long-term changes in other behavioural domains.

Green et al. (2010) ran a parent-mediated intervention known as the Preschool Autism Communication Trial. Social and communicative impairments were targeted, with the aim of training caregivers in improving responsiveness to their child. Video-feedback methods were employed in order to improve parent-child interactions, and a range of additional social strategies such as familiar repetitive language were encouraged. A total of 152 infants aged between 2 and 4 years were assigned either to a clinical intensive treatment programme, or a group receiving specialist treatment as would typically be offered. Intervention ran for a total of 13 months, consisting firstly of two intensive monthly appointments, followed by single monthly booster visits for the final 6 months. Whilst non-significant differences between pre- and post-intervention scores were reported in the standardised ADOS-G test, parental ratings of both language and communication scores were highly significant, as well as assessor rated parent-child interactions.

Interventions have shown relative success in demonstrating improvements across social and communicative domains. However, under the assumption that the brains of children are highly plastic earlier in development and at that point would have strayed less far from the typical range of developmental trajectories, earlier interventions could provide a better chance of redirecting less severe atypical brain development and therefore demonstrate a higher efficacy than interventions conducted later in development. Zwaigenbaum et al.

(2015) and Warren et al. (2011) suggest that early interventions are likely to be the most effective, and that interventions should combine tasks from a number of behavioural and developmental domains in order to optimise positive results across multiple cognitive and behavioural domains.

Two published studied to date have intervened with toddlers under 30 months of age. Dawson et al. (2010) incorporated a behavioural (in the form of ABA therapy) and developmental, relationship-based intervention in toddlers aged between 18 and 30 months in the Early Denver Start Model (EDSM) intervention. Individuals in the treatment condition received on average 20 hours of training per week from clinicians and at least 5 hours of parent delivery. EDSM techniques included the practise of shared engagement in realistic settings and a focus on both verbal and non-verbal communication. Significant improvements were reported in IQ, adaptive behaviour, and language abilities in the intervention group. Moreover, Dawson et al. (2012) reported findings from an EEG facial recognition study, which recruited a subset of individuals from the original EDSM intervention group, and community intervention group a year after intervention had ceased. Individuals from both intervention conditions exhibited normalised neural responses to images of faces that were in the same range as typical controls. However, those in the EDSM intervention group demonstrated faster neural recognition response when responding to images of human faces in comparison with objects (such as cars and toys). Dawson et al. (2012) suggest this is due to the focus of increasing attention to faces in the EDSM model condition.

A two-year follow up of individuals from the EDSM intervention group showed that positive effects were maintained across both behavioural and social domains (Estes et al., 2015), with

a lower level of core symptom severity and higher levels of social and communicative skills demonstrated at a group level. Taken together, these findings demonstrate progress in social and communicative domains as a direct result of early intervention trials; however, it also highlights the difficulties in attempting to ascertain the most successful methods for interventions. The fact that children in both the community and specialist intervention groups progressed and maintained learnt skills acquired from intervention suggest that it is perhaps not wholly about the specificity of the intervention. It instead indicates that duration and timing are of equal importance. Research suggests that earlier and more intensive intervention can provide longer lasting positive effects in children with ASD. One caveat, however, is that interventions have yet to be tested at both early and later stages of development. Thus, we have not been able to systematically compare early and late intervention effects using the same intervention methods. Anecdotal evidence is often relied upon to substantiate such claims, and it would greatly benefit our understanding of the effect of timing on intervention to compare results from intervention implemented at different developmental stages.

Green et al. (2015) recruited infants with a familial high-risk of developing ASD for the first intervention implemented within the first year of life. An advantageous aspect of using a high-risk population to study intervention effects is that very young infants can be employed. Infants aged between 7 and 10 months from the prospective, longitudinal British Autism Study of Infant Siblings (BASIS) were utilised. The intervention itself utilised a video feedback task for parents in order to help them to demonstrate better responses and sensitivity to everyday tasks involving social and communicative skills, and verbal and non-verbal communication (i-BASIS-VIPP). A total of six sessions were provided over 5 months, with six additional booster sessions. Successful effects of intervention were present in the levels

of infant attentiveness towards their caregiver, and improvements were exhibited in scores on the Autism Observation Scale for Infants (AOSI). A particularly interesting finding was that intervention effects appeared to display a generalised pattern of effects, which was spread across multiple cognitive and dyadic outcomes. In comparison, interventions with older infants tended to positively influence target outcomes. What can we conclude about this compensatory effect? Green et al. (2015) tentatively linked this finding to the social motivation hypothesis; specifically, that early intervention could provide the opportunity to positively alter brain growth trajectories, due to the higher levels of plasticity in the first 12 months of infancy (Webb, Jones, Kelly & Dawson, 2014).

The application of randomised control trials in intervention research has generated a number of findings that demonstrate the possibility of long-term gains in language, communication and social skills, across both cognitive and developmental domains. Furthermore, preliminary results from very early interventions have shown wider-spread benefits as opposed to task specific improvements seen in interventions implemented at a later point in development. The most effective intervention programmes lasted for a longer duration and were of a higher intensity in terms of the amount of sessions per child. Many researchers and clinicians also view direct parental training and involvement as an integral part of a successful intervention process. However, a number of methodological limitations mean that definitive conclusions are yet to be established from these studies. Firstly, small sample sizes make generalisability problematic. Many samples were self-chosen, and the vast majority of children involved in intervention trials were from affluent, middle class families. It must be questioned whether positive intervention effects would be as prominent in children who were not exposed to such constructive environments.

Moreover, a high level of variability is still observed between children post-intervention. Whilst some children successfully made gains across numerous cognitive domains, others failed to progress although they were exposed to the same interventions. It has been argued that introducing diagnosis-specific interventions could improve the effectiveness of research programs. Zwaigenbaum et al. (2015) argue that 'critical ingredients' from intervention procedures must be identified through varying components of a multi-discipline intervention program. A second method by which variability can be further understood is to characterise the behaviours of individuals over developmental time through the implementation of longitudinal, prospective studies. This approach will enable researchers to examine the complex interplay between cognitive, behavioural and biological mechanisms underlying heterogeneity, and would provide the opportunity not only to tailor specific strategies to individual needs, but also the potential to initiate intervention at the earliest time points by identifying the primary cognitive and behavioural atypicalities in ASD.

1.6 Variability

The phenotypic variability seen in ASD has led to difficulties in understanding the aetiology and mechanistic underpinnings of the disorder. As previously discussed, responses to intervention programmes and subsequent developmental outcomes are extremely variable between individuals with ASD. Heterogeneity spans a range of motor, social, language cognitive domains as well as IQ. This variability has, thus far, prevented consistent descriptions of the differences between those with ASD and typical populations, and as a result, clinical observation based on viewing actions and behaviours are still the primary tool for diagnosis. Levy and Epstein (2008) argue that the imprecision of behavioural phenotyping is the single most important contributing factor in failing to explain underlying biological and genetic factors in neurodevelopmental disorders.

More recently however, researchers are investigating variability directly. In the following section I describe genetic, neurobiological and clinical data that are attempting to address the issue of variability by elucidating the neurobiological, genetic or environmental causes that could account for the full constellation of behaviours seen in individuals with ASD. First, genetic findings are discussed, comparing the identification of common mutations and individual candidate genes. The question of whether autism can be defined as one or multiple disorders is addressed using findings from genetics, neuroimaging and biomarker research. Subgroups based on behavioural phenotyping to address the issue of variability are also discussed.

1.6.1 Genetic variability

Heterogeneity is striking across all explanatory levels of ASD, and the challenges in behavioural research are also reflected in the field of genetics. The risk for ASD is familial, with an overall estimated heritability level of at least 90% (Geschwind & State, 2015). Because of this, a large body of genetic research initially focused on identifying candidate genes with the potential to account for a large proportion of individuals with ASD that could directly relate to the behavioural phenotype. Research previously suggested that between 10% and 20% of all cases of autism are caused by individual candidate genes, an example of which is known as de novo mutations (Sebat et al., 2007). These mutations are rare sequence or copy number variants and appear spontaneously in individuals rather than being inherited from a parent. It has been estimated that around 14% of individuals with ASD carry a de novo mutation linked to the disorder (Sanders et al., 2012).

However, a second strand of genetic research suggests that much of the genetic risk for ASD is commonly found in the general population rather than originating from rare gene variants. The study of these genes, known as common gene variants has further elucidated the extent of genetic contributions to the risk of developing ASD. Common variants, usually variations or mutations, are present in at least 1% to 2% of the population and individually, exert only a small influence on determining atypicality (Constantino & Todd, 2005). However, common variations in the genetic code of an individual can add up to a substantial impact when combined, and significantly increase the likelihood of the occurrence of ASD (Buxbaum et al., 2014). Gaugler et al. (2014) reviewed over 500,000 variants in several thousand non-related individuals in order to identify the extent to which common variants contribute to autism risk overall. They estimated the contribution of common variants to be around 52%. In comparison, the effect of de novo or other rare mutations was just 3%. Similarly, Cook and Scherer (2008) reported that individual factors fail to explain more than 2% of the genetic risk of ASD. Sandin et al. (2014) also found a similar level of genetic risk. This, in principle, can explain much of the variation in multiplex families (where a person diagnosed with ASD has at least one first- or second-degree family member with an ASD diagnosis). It is possible that a number of first-degree or second-degree relatives are diagnosed with ASD, or that several family members display multiple autistic traits, but may not receive an ASD diagnosis.

To account for the large percentage of individuals where single genetic mutations are not likely to be the overriding factor for the onset of ASD, Elsabbagh and Johnson (2010) introduce two theories of "gene-dosage" models. The first is a cumulative model of development, in which numerous genetic and environmental risk factors could interact in an additive way, resulting in an atypical phenotype and variability. The second model explains

development of ASD as complex interactions between genetic and environmental processes during the earliest phases of development, when the brain is undergoing a period of optimal plasticity. It is argued that this process explains the varied trajectories exhibited in ASD. This theory is supported by an infant intervention study, which demonstrated wider cognitive and neural effects of intervention in younger individuals (see Green et al., 2015, above). However, the precise influences involved are yet to be identified.

1.6.2. Autism or autisms?

Given that ASD is currently defined on behavioural grounds, the question of whether autism is comprised of one or multiple distinct disorders is about elucidating the underlying genetic and biological causes of behaviours specific to ASD. Until recently, the widely held view among researchers was that behavioural atypicalities associated with ASD occur more frequently together than would be expected at a chance level; therefore, it would be expected that a single causal pathway may be able to explain the manifestation of these symptoms (Boucher, 2011). However, the heterogeneous nature of the disorder in terms of both development and outcome has led to a divergence amongst researchers, with the suggestion from some that multiple causal pathways to ASD could be identified (Ronald et al., 2006; Geschwind & Levitt, 2007). When behaviourally defining a disorder such as ASD, one must ask how we can be sure that we are studying a single disorder with a single underlying mechanistic cause, and not incorporating a number of distinct syndromes that exhibit similar behavioural markers. Currently there is no consensus over which is more likely to be correct.

The identification of multiple developmental trajectories in ASD has led the proposal that autism occurs on a mechanistic continuum (Zwaigenbaum et al., 2009) and that ASD profiles defined by a loss of skills or a plateau in abilities in early development may represent the

extreme end of a continuum of trajectories (Ozonoff et al., 2011; Ozonoff et al., 2010). This is also supported by findings from the Over-Pruning hypothesis, a computational account of ASD, which identified multiple atypical developmental trajectories as a result of a single pathological mechanism interacting with individual differences as a result of population variation (Thomas, Knowland & Karmiloff-Smith, 2011; Thomas, Davis, Knowland, Karmiloff-Smith & Charman, 2015) (see Chapter 4 for detailed findings and below for further details about this account and computational methods).

Alternatively, others have supported accounts that propose multiple mechanistic causal pathways in ASD. Happé and Ronald (2008) argue that progress in our understanding of heterogeneity has remained somewhat static because of the assumption that one underlying cause can explain all social and non-social symptoms that define ASD. Utilising results from population-based genetic studies (Ronald et al., 2005, 2006), the authors demonstrate weak correlations between the core traits of repetitive interests and social and communication, and instead propose the 'fractionable triad hypothesis'. The authors suggest that the phenotypic components of the triad of impairments in ASD could arise from distinct genetic clusters. As well as only identifying weak correlations between core behavioural markers in ASD, they also report that many children in the sample only exhibited difficulties in one domain. For example, 59% of children with social impairments did not display impairments across any other domains. This is supported by a number of studies that have also identified significant distinctions between core symptoms of ASD (Van Lang et al., 2006; DiLalla & Rogers, 1994). One point to make here is that the fractionable triad hypothesis utilised findings from children aged between 7 and 10 years. In order to fully understand the extent to which cognitive functions are fractionated, it would be necessary to evaluate the correlations between behavioural abilities at the earliest possible time points in development. Utilising

high-risk infant populations to track changes across development would identify the progression of behavioural atypicalities and could disentangle causal mechanisms and secondary symptoms.

When used alongside clinical research, neuroimaging studies have contributed to the delineation of developmental trajectories within ASD (Mahajan & Mostofsky, 2015). In this thesis, of particular interest is research that attempts to identify behavioural and anatomical differences between those with ASD and typically developing individuals. One such study proposed a general method for identifying and quantifying heterogeneity in the ASD population, using the population characterisation of heterogeneity (PUNCH) technique (Tunc et al., 2014). PUNCH uses a decision level fusion approach; here, results from 370 individuals with ASD were compared using nine tests measuring 50 traits associated with ASD (including social responsiveness, anxiety, language and communication). Results from the evaluations were combined in order to construct a single severity score for each child. The authors argue that samples could then be divided up in novel ways based on severity levels ascertained from multiple tests scores, providing an opportunity to identify subgroup characteristics that were previously undetected. Scores from each test domain were weighted according to how well a score distinguished typically developing individuals and individuals with ASD. A single score of severity was created for each individual based on performance percentiles compared across all individuals. A high severity score meant that those individuals exhibited more traits that differentiated them from the typically developing group. Scores were used to analyse brain images, comparing typically developing and ASD individuals. Severity was found to moderate group differences; specifically, brain differences between individuals with ASD and controls intensified with autism severity. When individuals with low PUNCH scores (or low severity) were removed from the analysis, the

authors identified more differences in comparison with assessing the ASD group as a whole. It is argued that future studies could implement similar statistical techniques, with the aim of identifying trends that might otherwise be concealed by those with milder forms of the ASD.

The identification of biomarkers is a second method that has been utilised to understand the mechanistic underpinnings of heterogeneity in ASD. Whilst clinical observations have been invaluable in characterising impairments in ASD, some argue that the progression of our understanding of the disorder will not continue unless we are able to identify more objective ways to track social deficits (Ratajczak, 2011). The aim of biomarker research is to pinpoint specific biological differences in individuals with ASD, which in turn may facilitate more precise diagnostic tests for identification. Much of the biomarker research in ASD has focused on levels of specific neurotransmitters in those who have already received a diagnosis (e.g. Auyeung et al., 2009; Pardo & Eberhart, 2007).

A current example that incorporates both biological and behavioural measures in their search is the Autism Biomarkers Consortium for Clinical Trials (www.asdbiomarkers.org), which has the aim of identifying reliable measures of social functioning. Neurological methods such as EEG and eye tracking will be used in conjunction with data from behavioural test batteries in order to further understand the interactions of behaviour, brain activity and attention. The authors argue that the large sample of 200 children could facilitate more sensitive detection of EEG based biomarkers, which in turn could enable clinicians to predict responses to specific interventions more reliably. Thus far, very few biomarkers that have been identified in autism have successfully translated into useful clinical applications. This is partly due to the changing phenotypic manifestations throughout infancy and early childhood (Walsh, Elsabbagh, Bolton & Singh, 2011). Whilst studies such as this are progressing our

understanding of brain differences in children with ASD, the fact that children are recruited at three years of age makes it difficult to distinguish between cause and effect when looking at significant findings without having knowledge from infancy and early development.

In terms of using biomarkers as predictors before overt behaviours arise, a small number of studies have utilised high-risk infant siblings to identify neural atypicalities (Zwaigenbaum, et al., 2009; Gliga, Bedford, Charman & Johnson, 2015) (see Chapter 3)). However, biomarkers identified thus far have not been robust enough to be meaningful at the level of the individual.

1.6.3. Subgroups based on developmental trajectories

Clinically defined sub-categories of ASD have been proposed as a way of accounting for the heterogeneity demonstrated between individuals with ASD in terms of the differences in the manifestation of symptoms. One possibility is that the identification of definitive, internally homogeneous, behaviourally defined subgroups within ASD could support the exploration of genes and biomarkers and reduce overall heterogeneity.

In the first of such studies, two subgroups were proposed based on atypicalities observed across developmental trajectories. Lord, Shulman and DiLavore (2004) utilised longitudinal data consisting of both interviews from parents whose children were referred for a possible ASD diagnosis, and standardised test batteries. Data were collected at 2, 3 and 5 years, with particular emphasis on regressive symptoms and developmental delay. One predominant pattern was typified by early atypicalities in social and communicative domains that developed across the first 3 years. The second pattern, known as the regressive subtype, was characterised by a phase of seemingly typical development followed by a phase of clinical

'regression' in this case in the form of word loss. A total of 25% of individuals reportedly followed this trajectory. Several other retrospective studies have reported similar findings, identifying both early onset (Luyster et al., 2005; Werner, Dawson, Osterling & Dinno, 2000) and regressive subgroups.

Research from prospective studies, however, has revealed a more complex picture of development in ASD. The implementation of prospective, longitudinal studies has allowed for advancements in the identification of subgroups through the use of sophisticated and precise behavioural and cognitive measures (see Chapter 2 for a review of measures). As previously discussed, retrospective research, though valuable, is prone to the effects of bias and misremembering. Ozonoff et al. (2010) conducted a prospective infant sibling study to address previous claims about regression and to further elucidate subgroups within ASD. Twenty-five infants who were later diagnosed with ASD were assessed using a number of developmental measures and behavioural coding at multiple time points between 6 and 36 months. In addition, findings from retrospective parental reports were included in analyses. Crucially, whilst cognitive and behavioural measures demonstrated a decline in developmental trajectories across multiple domains between 5 and 18 months in the majority of infants, 83% of parents reported no loss of skills during this time. This is a key example of the inaccurate representations of development that can be captured in retrospective designs.

Ozonoff et al. (2010) (also Ozonoff, Heung, Byrd, Hansen & Hertz-Picciotto, 2008) suggest that a traditional dichotomous categorisation fails to encompass all of the ways that ASD develops (alternatively, it could be that individuals with multiplex ASD, necessary for at-risk studies cannot be categorised in the same subgroups). They suggested that as many as four developmental trajectories or subgroups in ASD may exist. The majority of prospective

studies have reached a general consensus that an early onset subtype is present in some individuals (Ozonoff et al., 2010; Landa & Garrett-Mayer, 2006; Landa, Holman & Garrett-Mayer, 2007). Ozonoff et al. suggested that additionally, individuals with ASD may demonstrate developmental regression, a combination of early onset and regressive behaviours, or exhibit typically developing social and communicative functioning in the first 6 to 12 months, followed by an inability to learn novel skills or progress in a typical fashion after this point. This is described as "pseudo regression" (Klin et al., 2004) or a developmental plateau. Similarly, Landa, Gross, Stuart and Bauman (2012) identified early onset (identified by 14 months) and late onset (identified after 14 months) subgroups, as well as a decline in development and plateauing in a number of individuals. The identification of these subgroups is further supported by other prospective studies to date (Zwaigenbaum, Bryson, Rogers & Szatmari, 2005). Importantly, change in variability and severity between subgroups across time can also inform us as to whether ASD is comprised of one or multiple disorders. Thus far, some support has been gathered for the idea that single mechanistic cause underlies different symptom onsets in ASD because of the lack of divergence between regressive and non-regressive groups later in development (Heung, Ozonoff & Byrd, 2011). However, additional research is necessary in order to validate such claims.

It can be questioned whether any of the subgroups identified thus far have provided a strong case for a neurobiologically robust and distinct subtype of ASD. However, it is still possible that biological subtypes are associated with specific and individual cognitive or behavioural phenotypes. Nevertheless, Charman et al. (2011) propose that the existence of subgroups based on patterns of atypicality in cognitive domains may be essential for optimising intervention strategies. This would proceed firstly by aiming to improve in areas of difficulty either through practise or by finding alternative learning routes to understand particular tasks;

and secondly, strengths identified in particular areas of cognition can also be utilised to overcome other areas of weakness. Furthermore, Landa et al. (2013) suggest that fully understanding trajectories provides the potential to clarify the role of brain mechanisms and the extent to which typical synaptic plasticity is disturbed in individual with ASD. They proposed that this interruption (through both genetic and environmental factors) could disrupt "brain readiness" for the connection of learning mechanisms and neurobiological processes, which would in turn affect outcome and severity of the disorder itself.

1.7 Computational models as a tool for understanding development

Computational models have more recently served as a complementary tool to further improve our understanding of developmental disorders in terms of the underlying mechanistic causes and differences in developmental outcomes. It has been argued that developing appropriate models of neurodevelopmental disorders will help to move from methods that simply observe developmental changes over time, to more dynamic accounts which elucidate the 'well specified' processes underlying these emerging behaviours (Mareschal & Thomas, 2007).

Computational models can benefit the study of neurodevelopmental disorders in numerous ways, but in order to do so, Mareschal and Thomas (2007) suggest a number of measures to be taken within this framework. For example, they advise that a model must be conceived on the basis of available data from the domain of cognitive development. In order to utilise modelling and successfully contribute to research in a complex developmental disorder such as ASD, researchers must have a full understanding of the current frameworks, theories and debates within that field. Furthermore, processes that are used and suggested within the model should be relevant to mechanisms in other domains. A theory at a computational level

for example, should contain mechanisms that are compatible with neural, genetic or cognitive levels of processing in order to generalise findings efficiently.

One aspect of modelling that has both positive and negative implications is simplification. When building a cognitive model for the study of developmental disorders, experimental research is often the basis for such work. This ensures that neurocomputational parameters and training sets are developed in the most realistic form, and remain useful and generalisable. The idea when implementing findings from previous research is to build a system that focuses on the key mechanistic elements, and removes unnecessary aspects (Thomas, Baughman, Karaminis & Addyman, 2011). If a model can successfully achieve this, then it enables a greater level of precision and clarity, and importantly, can examine the viability of multiple theoretical claims that in empirical studies would be impossible or unethical. For example, a computational model intervening on an atypical learning system could examine the same individual networks under multiple different experimental conditions, in order to compare the effectiveness of different types of intervention. Furthermore, a huge number of individual networks can be run consecutively, eliminating the problems present in clinical research where it is costly and difficult to recruit high-risk infant siblings, for example.

There are universal limitations when using computational models. The simplification of cognitive data can lead to problems in relating abstract mechanisms back to potential neural or cognitive levels. Whilst a neural network approach has a foundation in neural constructs, it is not possible to directly map neurocomputational parameters onto neural functions. One must remember that computational processes will, to some extent, inevitably remain an abstract form. Furthermore, the tasks learnt by networks also often take an abstract form, and

the learning problem is not always directly relevant to concrete instances of behaviour. Whilst models serve to establish the validity of theoretical hypotheses, they cannot determine whether there is truth behind such proposals (Mareschal & Thomas, 2007). However, if used with consideration, computational modelling can make a vital contribution to the understanding of developmental disorders and cognitive development. A number of computational models have been proposed suggesting causal mechanisms in ASD (see Chapter 3 for a full overview).

1.7.1 The Over-Pruning account of ASD

Part of this thesis expands on findings from a computational developmental model of ASD. While a broader theory of the origin of ASD, the Over-Pruning hypothesis (Thomas, Knowland & Karmiloff-Smith, 2011) originated as a neurocomputational model aiming to understand the underlying causal mechanisms of regressive subtypes within ASD. Further investigation however, has led to a theory that encompasses the heterogeneity in both timing and expression of ASD (Thomas, Davis, Karmiloff Smith, Knowland, & Charman, 2015). Population modelling in an artificial neural network model was implemented. This is a technique whereby the development of a large number of individual networks is simulated. Individual differences are produced in each network by varying the unique learning environment to which a single network is exposed through its lifespan, and varying learning abilities through the learning parameters. Networks were tested on an abstract training set for the duration of development, with varying levels of task difficulty (see Chapter 4 for a full overview of the model).

The Over-Pruning account proposes that ASD arises from an overaggressive synaptic pruning mechanism. Across the first two years of infancy, the trajectory of typical brain development

is typified by robust growth and increases in connectivity, followed by a period of pruning in order to remove unused and redundant connections (Huttenlocher, 2002). The hypothesis here is that in individuals with ASD, an aggressive form of this process is responsible for regression. Furthermore, this synaptic pruning occurs within a typical timeframe across development, supporting findings from prospective data that the profiles of infants who go on to develop ASD present a seemingly typical developmental trajectory across the first 6 months of life (Rozga et al., 2011; Ozonoff et al., 2010; Damiano et al., 2012; Clifford et al., 2013).

A number of conclusions have been drawn from the model. First, atypicality (as opposed to regression) is explained by a combination of an exaggerated pruning mechanism and individual differences within a population. When individual differences in the rate of learning, rate of pruning, capacity to learn, and previously learnt information interact with the pruning process, variation in developmental trajectories and outcomes occur. Thomas et al. (2015) argue that population-wide individual differences can serve as protective or risk factors. For example, at the point of pruning onset, early experiences could serve to protect those domains that have been strengthened, which could explain the heterogeneous profiles observed in ASD. Moreover, simulations demonstrated that a small number of individuals were susceptible to pruning in the low-risk population where the severity of pruning was in the normal range. This suggests two pathways to regression; either due to an extremely high pruning threshold or through an unlucky combination of individual differences caused by risk factors and a pruning threshold that while low are in the normal range.

This computational account identifies potential underlying mechanistic processes in ASD, and is novel in that it emphasises the developmental process underlying atypical cognition.

To my knowledge, this is the first developmental computational model (see Chapter 4 for a review of computational models of ASD) using simulated individuals to identify the influence of inherent and external differences on outcome and performance. It is vital that cognitive theories of ASD incorporate developmental processes into hypotheses. Whilst it must be said that this model is preliminary and based on abstract representations rather than anchored in specific behaviours, development is seen as a dynamic interactive process (Karmiloff-Smith, 2009). I would argue that the use of modular, static theories of ASD will not be able to fully explain the disorder, or contribute to understanding the core issue of heterogeneity.

Within this thesis, I will clarify some of the conclusions from the Over-Pruning model through the comparison with longitudinal prospective data, and the use of further computational models. First, variability and subgrouping is explored, and the impact of early versus later manifestations of the disorder. Findings from the model suggest that earlier atypicalities may be associated with poorer outcome, due to the fact that connectivity will be more robust later in development as additional time has allowed for more experience-dependent strengthening. Furthermore, during typical development, sensory areas of the brain are the first to undergo synaptic pruning (Huttenlocher & Dabholkar, 1997). Thomas et al. (2015) predicted that the first primary emerging symptoms in ASD should therefore be identified in sensory and motor domains as opposed to social. This question is also investigated in the clinical population. Third, it was claimed that following the behavioural manifestations of the disorder, intervention cannot be used to normalise a system, but instead can compensate for the damage that has already occurred. Therefore, if intervention is used in a protective manner (against pruning) then it should commence as early as possible. Furthermore, since pruning is widespread across all brain areas, it is argued that intervention

should target multiple cognitive and behavioural domains in order to maximise compensatory mechanisms. Based on these claims, I utilise results from targeted interventions on three model populations. I compare attempts at normalisation and compensation and investigate the effects of timing and intervention.

1.8 Discussion

There are a vast number of theories of ASD, the majority of which follow a modular, static account. In comparison, others have proposed developmentally driven theories, which posit that ASD is a product of complex interactions between multiple levels of explanation, and changes across time due to brain plasticity in infancy. One of the greatest challenges in elucidating underlying causal mechanisms is the heterogeneity amongst individuals with the disorder. Emerging findings from prospective, longitudinal studies of development have provided some evidence of subgroups and positive effects of intervention; however, more research looking at variability and response to intervention across developmental trajectories is necessary to further understand the subtle changes in early development. In Chapter 2 of this thesis, I look at between- and within-subject variability using data from at-risk siblings from the first phase of data collection from the BASIS longitudinal prospective study of autism. I also evaluate hypotheses from the Over-Pruning computational model of ASD. I then utilise 14-month data from the first cohort of the BASIS study to identify behavioural and environmental measures that can be used as a cumulative risk index to predict outcome at 24 and 36 months. I then use the same index to predict outcomes of individuals from a second independent dataset at 24 months, the outcome data at the time of writing (Chapter 3). Chapters 4 and 5 consider data from our computational account of ASD. In Chapter 4, I compare two novel populations from the Over-Pruning model and identify subgroupings from developmental trajectories in regressive and atypical individuals. I use statistical

models to regression models to identify the differences in underlying mechanisms that differentiate the groups. Chapter 5 investigates recovery and the effect of intervention in the modelling populations from Chapter 4. I examine the effects of two types of intervention, one that attempts to compensate, and one that attempts to normalise the learning system, where compensation and normalisation are given specific machine-learning based definitions. I compare across populations and previously identified subgroups. In the discussion (Chapter 6) I relate current findings to research introduced in the current Chapter, and link computational and empirical results. I consider the limitations of this thesis and discuss future directions.

The overall aim of this thesis is to combine the analyses of clinical empirical data and data from a computational model of ASD in order to elucidate the influences on developmental trajectories of autism, to further understand variability manifestation and outcome, and to trace recovery and understand response to intervention. I focus on a number of central questions:

For the clinical data:

1) What is the relationship between outcome and variability in the high-risk groups? How do the patterns of variability differ between domains and outcome groups across 36 months? I used behavioural measures from two standardised tests to investigate whether more uneven cognitive profiles in infancy were indicative of later clinical outcome. I then compared heterogeneity levels in the ASD and low-risk groups.

2) To what extent can behavioural risk factors be identified at 7 or 14 months and used to predict outcome at 36 months? Here, risk factors identified at 14 months from behavioural and environmental measures were included in statistical regression models in order to predict diagnostic outcome at 24 months. The same statistical model was then tested on data from a second group of infants, followed from 5-36 months (known as Phase 2 infants) to investigate whether outcome could be predicted at 24-months in an independent sample.

For the computational modelling data:

3) Can subgroups be identified in the computational populations? What are the mechanistic processes accounting for group differences? I identified groups based on recovery rates after regression and outcome levels in atypical, non-regressive groups, using statistical models to pinpoint the relevant neurocomputational parameters in each subgroup.

4) Can intervention be successfully implemented in atypical networks? I used individual networks from atypical subgroups to investigate the effects of two types of behavioural intervention: First, attempts to normalise the system, second by the use of compensatory mechanisms. Intervention groups were then compared.

Chapter 2

2.1 Introduction

As previously mentioned, our understanding of ASD has progressed considerably through the implementation of sophisticated study designs such as prospective longitudinal methods, advancements in the utilisation of neuroscientific techniques such as EEG and eye tracking and initial findings from computational models aimed at understanding the processes underlying ASD. However, one area that still poses considerable challenges in elucidating the underlying neural mechanisms is the heterogeneity or variability observed in ASD. Variability is present in the phenotypic manifestation of the disorder, occurring both in terms of the strengths and weaknesses displayed in an individual's cognitive profile and in the stability over the course of development and in later years (Charman et al, 2015). In the current chapter, I investigate the effects of both within (intra) and between (inter) subject variability by utilising data from the BASIS prospective, longitudinal study of high-risk infants.

In order to reduce the effects of heterogeneity at a group level, attempts have been made to categorise individuals with ASD into more homogenous groupings. The current chapter presents the status of such subgroups. I also discuss findings regarding the stability of diagnosis of ASD across childhood and adolescence and more specifically, across the first three years of development. Much of the prospective research into early atypicalities in ASD has utilised group mean ability scores. Here I measure variability in ability scores between outcome groups, specifically in social, language and motor domains. I then consider

variability between developmental measures at four time points and identify whether variability is significantly higher in individuals with ASD compared to high-risk individuals without an ASD diagnosis and low-risk infants. Based on reports of variability from prospective studies (see below), I expect variability to increase over the first 3 years of infancy as the behavioural profile of ASD emergences. I examine the extent to which an uneven cognitive profile (as measured by an individual's variability across domains) was correlated with cognitive ability levels, and whether this was predictive of outcome. In order to compare heterogeneity between time points, variability is compared between 7 and 14 months, 14 and 24 months, and 24 and 36 months in all behavioural domains to assess the association between intra-subject variability and later ability levels in behavioural domains.

2.2 Within-subject variability and stability of diagnosis

The focus of ASD research has shifted towards theories that encompass heterogeneity whilst attempting to elucidate the underlying influences on outcome and development (Milne et al., 2006; Milne 2011; White et al., 2009). One approach has been to evaluate variability within individuals over time, and a key area of research focuses on the stability of diagnoses later in development. Several studies have reported a high degree of diagnostic stability in children aged between 2 and 3 years at the time of initial diagnosis, with figures in the range of 84% to 100% in clinically referred samples (Charwarska et al., 2009; Charwarska, Klin, Paul & Volmar, 2007; Woolfenden et al., 2011). Furthermore, a number of studies have followed infants through early childhood up to 7 years of age (Shepherd et al., 2016; Charman et al., 2005) and 9 years of age (Lord et al., 2006), and findings have supported claims of moderate levels of stability within individuals diagnosed with ASD. However, it has been argued that the stability of diagnosis is lower in individuals diagnosed with broader ASD profiles. Hedvall et al. (2013) measured cognitive abilities in 208 pre-school children with ASD and

individuals with broader ASD symptoms over two years. They reported considerable changes in developmental profiles during this time relating to persistence of diagnoses and intellectual outcomes. Moreover, Ozonoff et al. (2015) assessed the stability of early autism diagnoses. Using data from the Baby Siblings Research Consortium, the authors evaluated whether children could be accurately diagnosed with ASD before 36 months of age. The authors utilised data from 418 at-risk siblings, who were assessed at 18, 24 and 36 months. Whilst the stability at both 18 and 24 months was reasonably high (93% and 82% respectively), almost 50% of the children with ASD were not identified at 24 months, and did not receive a clinical diagnosis until 36 months of age. The authors suggest that screening procedures for ASD should be conducted at multiple time points, particularly in high-risk children, until 36 months, where diagnosis is more stable.

Two recent studies have also assessed diagnostic stability by questioning the established view that ASD is a lifelong developmental condition. One retrospective study (Fein et al., 2013) and one prospective study (Anderson, Liang & Lord, 2014) reported a reduction of ASD symptoms over time in a small number of individuals (a total of 34 individuals, and 8 individuals from a sample of 85 respectively), such that they no longer met the criteria for ASD (see Chapter 5 for a full discussion). The small sample sizes and the use of retrospective data somewhat limit the generalisability of these findings. However, they have led to the consideration that a small number of individuals diagnosed with ASD in infancy may 'lose' their diagnosis later in development. If findings can be replicated in other samples, it will be essential to elucidate both internal and external factors (IQ and early ability levels, the effects of early intervention) that differentiate individuals displaying 'optimal outcomes'.

The variability in stability of diagnoses is an example of the difficulties presented when identifying and treating a behaviourally defined disorder. The current consensus is that an ASD diagnosis remains moderately stable across development. However, lower stability levels have been reported in individuals exhibiting a broader range of symptoms, and a small number of individuals have seemingly 'lost' their ASD diagnosis by adolescence.

2.3 Subtyping in ASD

It has been argued that the identification of accurately defined subgroups could benefit the search for genetic and biological atypicalities by reducing the heterogeneity that is present when individuals with ASD are viewed as a single group (Charman et al., 2011). To date, a handful of studies have attempted to subgroup cognitive profiles in ASD, which are discussed below.

One argument for cognitive subtyping in ASD has been fuelled primarily by the proposal that identifying individuals with specific cognitive atypicalities could provide the opportunity for individually focused interventions. If a homogenous group of individuals with ASD could be identified based on areas of cognitive strengths, these skills could be developed to overcome weaknesses in other cognitive domains (Charman, 2015). Jones et al. (2009) identified one subgroup based on auditory discrimination abilities in 72 adolescents with ASD. The study focused on differences in sound intensity, frequency and duration using paired sounds. When assessed as a single group, no significant differences in discriminatory abilities were identified in the ASD group when compared with typically developing controls. However, a subgroup containing 20% of individuals from the ASD population exhibited elevated levels of frequency discrimination in comparison with non-ASD groups. Furthermore, the ASD subgroup that demonstrated enhanced discriminatory abilities also exhibited comparable

cognitive profiles. This profile included average intellectual ability and delays in attaining language milestones during early development. This pattern of findings led to the authors' suggestion of a putative specific phenotype in ASD and has since been replicated (Bonnel et al., 2010). The authors reported enhanced pitch discrimination for simple tones in individuals with ASD, but not in those with a diagnosis of broader ASD.

Several studies have attempted to identify subgroups in ASD using cluster analysis to characterise distinct clinical profiles. Wiggins et al. (2012) utilised a sample of 186 toddlers with a clinically defined ASD diagnosis, and scores from the Childhood Autism Rating Scale (Schopler & Reichler, 1980), which included measures for social, communicative and intellectual skills, as well as repetitive and restricted interests and behaviours. Subsequently, three subgroups, or clusters of ASD, 'mild' 'moderate' and 'severe' were identified based on symptom severity. In particular, it was not the presence of these behaviours but the rate and intensity at which they occurred. Specifically, 76% of variance that differentiated individuals in different ASD subgroups was accounted for by differences in social and communication skills. The authors propose that these abilities are particularly relevant in the classification of young children with ASD. Moreover, Wiggins and colleagues question whether it is more appropriate to classify individuals with ASD categorically or by adopting a dimensional approach. They argue that their results support a dimensional approach when diagnosing ASD, as subgroups were distinguished by level of social, communication, and intellectual abilities rather than distinct symptom profiles. Furthermore, Georgiades et al. (2013) proposed that two symptom dimensions; social-communication deficits and fixated interests and repetitive behaviours can be used to classify children with ASD into three distinct, homogenous subgroups. The authors analysed data from 391 children who had been diagnosed with ASD by three years of age. Specifically, data from the Autism Diagnostic

Interview-Revised that related to the two dimensions of interest were analysed. Three clusters of subgroup were identified based on the severity gradients of the specific clinical symptoms.

Kim, Macari, Koller and Charwaska (2016) argue that features outside of autistic severity should also be included in the identification of reliable homogenous subgroups. In their study, 100 toddlers were first evaluated for ASD between the ages of 14 and 27 months. A cluster analysis utilised scores from measures of ASD symptoms (social affect and repetitive and restrictive interests from the Autism Diagnostic Observation Schedule) and general cognitive measures of motor, social, communicative and daily living functioning from Mullen and Vineland tests. As with Wiggins et al. (2011) and Georgiades et al. (2013), high and low functioning groups were identified. Here, however, four groups were identified. The highest and lowest functioning groups constituted 53% of the sample, whilst the remaining individuals were categorised into one of two intermediate groups comprised of combinations of relative strengths and weaknesses. Interestingly, infants were followed up a year later in order to measure diagnostic stability. Individuals who were grouped in intermediate or severe clusters demonstrated a high stability (93%), whereas for individuals in the lowest severity cluster, stability was around 85%, with 6% of individuals no longer meeting the criteria for ASD. During this time however, intensive intervention treatments were on going, and it is not known whether these individuals improved considerably after intervention or whether symptoms presented at two years were similar to autism but caused by other underlying problems. Nevertheless, these studies suggest that it could be beneficial to consider the determinants of homogenous developmental trajectories in ASD across the first three years of infancy.

A small number of studies have attempted to identify subgroups by comparing individuals with ASD and ADHD, a frequently co-occurring disorder (Van der Meer et al., 2012). Here the aim was to examine the specificity of cognitive profiles and elucidate whether ASD-ADHD comorbidity subgroups existed by comparing individuals with ASD, ADHD and those with a diagnosis of both ASD and ADHD. A latent class analysis was used to identify five classes by comparing cognitive skills including motor speed and variability, attention and executive functioning. Both ADHD and ASD specific classes were identified and each group displayed specific patterns of cognition. However, two classes showed comorbidity between the disorders, and in each case one domain was more prominent than the other. In these individuals the cognitive profile was characteristic of each disorder, however, a greater severity level was demonstrated across cognitive domains compared to those in a single disorder group.

2.4 Developmental trajectories

Researchers have also used prospective studies to investigate differences in developmental trajectories over the first three years of life. In such cases, trajectory modelling has examined differences between the manifestation of symptoms across the first two years of development and the impact on outcome at the time of diagnosis. A number of trajectory–based subgroups have been proposed based on the timing of symptom onset (Landa, Gross, Stuart & Bauman, 2012; Landa, Gross Stuart & Faherty, 2013) or by identifying divergence between individuals demonstrating regression, developmental plateau, or early or late onset ASD (Ozonoff et al., 2010, 2011).

If it were possible to identify homogenous subgroups based on early cognitive and behavioural profiles, it would provide an opportunity for personalised interventions based on

cognitive strengths, where positive outcomes could be gained by nurturing areas of ability. It is feasible that these subgroups could be utilised to overcome difficulties in other domains and that building on cognitive strengths could have 'knock on' effects on others. One example of this is in dyslexia research in educational settings, where some teaching methods focus on identifying each child's strengths in order to circumvent weaknesses. This in turn has enabled children to learn more successfully by compensating for interferences in learning (Hale et al., 2016; Levy, 2008). However, in order to substantiate findings in ASD research, replications will be crucial in elucidating whether putative subgroups can be identified in other samples.

2.5 Variability as a measure of outcome

The majority of experimental findings in ASD compare mean task performance levels between individuals with ASD and a control group. However, it can be argued that this paradigm underestimates individual variability within groups. By focusing on the direction of mean differences between groups, some of the individual variability is ultimately masked. This is demonstrated in a number of studies that have reported higher levels of heterogeneity in ASD groups. Charman et al. (2005) reported that scores of individuals with ASD became more diverse over time in multiple cognitive measures. Furthermore, Landa et al. (2012) reported that heterogeneity increased with age across ASD groups in the Mullen early learning composite.

ASD is a disorder defined on the basis of behavioural impairments. Therefore, one theoretical question is whether individuals with an ASD diagnosis share the same underlying biological deficits, or whether individuals are grouped together because their behavioural impairments are similar, but in fact do not share the same mechanistic cause. Thomas (2003)

used computational modelling techniques to explore whether variability identified in behaviourally defined developmental disorders stems from individual variation within a group, and therefore demonstrates a single cognitive cause, or whether variability has multiple mechanistic causes whereby each sub-group demonstrates individual variability. Two disorder groups were created in the model; one group with a homogenous underlying cause, one with a heterogeneous cause. The two were compared in order to see whether groups could be predicted based on behavioural grounds. First, significant differences were identified between some groups in mean performance scores over a number of behavioural metrics. Conversely, other groups were not differentiated through performance scores. However, some were differentiated when the *variability* of performance measures was observed in each group over time.

Thus far, studies identifying homogenous subgroups and measuring the stability of diagnosis and assessing developmental trajectories in early infancy have provided insight into some aspects of variability in ASD. However, prospective research has not focused on utilising variability in ASD as a measure of development itself. A number of studies have reported increasing levels of heterogeneity within ASD groups over time. Furthermore, Thomas (2003) proposed that the variability of measures across time might be indicative of group outcome.

In the current chapter, I assess between-group variability and variability both within and across measures and across development in ASD using cognitive and behavioural data from the BASIS prospective study. Variability levels are compared between low-risk and high-risk subgroups across cognitive and behavioural domains. I then consider the extent to which an uneven cognitive profile could be predictive of lower ability scores and diagnostic

outcome at 36 months by comparing variability between cognitive measures. I examine within-subject variability by comparing variability between time points within single cognitive domains. Where variability levels were elevated in individuals with ASD, I measured the extent to which this was associated with cognitive measures and diagnostic outcomes at 36 months.

2.6 Method

2.6.1 Sample

Recruitment and background data on contributing families were made available for this study through the British Autism Study of Infant Siblings (BASIS), a collaborative research network utilising infants with a higher, familial risk of autism (www.basisnetwork.org.uk). In total, 104 infants were recruited, and attended up to four assessments at around 6 months, 14 months, 24 months and 36 months. Of these infants, 54 were categorised as being at a high-risk of developing ASD. High-risk infants were recruited on the basis that they had an older sibling with a confirmed clinical diagnosis of ASD. Diagnoses were confirmed by two clinicians based on information from Development and Wellbeing Assessment (DAWBA; Goodman et al. 2000) and the Social Communication Questionnaire (SCQ; Rutter et al. 2003). At the time of testing, none of the infants were diagnosed with a developmental or medical condition and medical histories showed no significant concerns in siblings or family members, therefore no exclusions were made on this basis. Fifty low-risk infants were recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. Low-risk controls were full-term (gestational ages between 37 to 42 weeks; 3 infants born between 32 to 36 weeks). Inclusion criteria included the stipulation that there was no ASD in first-degree family members. All infants had at least one older sibling (in three cases, only half-siblings). The SCQ was used to confirm that ASD was absent in all

siblings, and no children were excluded. In the current chapter, data from all time points were utilised

2.6.2 Measures

Infant developmental level

The Mullen Scale of Early Learning (MSEL; Mullen, 1995) is a standardised test used to measure cognitive functioning in children from birth to 68 months. The MSEL is often used to determine cases of atypicality, and when assessing an individual's need for early intervention (Dumont, Cruse Alfonso & Levine, 2000). The MSEL consists of five subdomains: visual reception, expressive language, receptive language, fine motor and gross motor. The visual reception scale analyses visual processing and visual memory. Tasks include recalling picture items, manipulating objects and fixation tasks. The expressive language sub-domain measures language formation and production. Infants are tested on spontaneous expressions and forms of verbal responsiveness. The receptive language scale assesses language comprehension. Children are assessed on their understanding of phrases such as "give it to me" and a combination of phrases with and without gestures are used. The fine motor sub-domain measures an individual's ability to utilise and make small, controlled movements, pointing gestures and to use coordination and control to manipulate small object. Tasks include asking the child to make hand movements such as a pincer grip and pointing motions, stacking blocks, drawing and stringing beads. The gross motor scale assesses larger elements of motor control such as standing and walking. Tests include asking the child to sit up right, crawl or walk towards an end point, and for older children using stairs. Scores can be combined across all five domains to create the early learning composite (ELC). However, in this chapter the five sub-domains were assessed individually. Table 2.1 provides the raw mean and standard deviation scores for low-risk and high-risk infant groups as a whole as

well as high-risk outcome groups based on 36-month classifications. Infants were either from the low-risk or high-risk groups. High-risk infants were then categorised as high-risk non-ASD (which incorporated all high-risk individuals who did not meet the diagnostic criteria for ASD), and a high-risk ASD group (see below). Differences in the number of individuals in Mullen and Vineland descriptive statistics are due to infants being unable to complete some of the assessments. Out of the 54 high-risk infants recruited, only 53 were assessed at the final 36-month visit.

Table 2.1 *Mean and Standard Deviation Mullen raw scores for low-risk, high-risk, high-risk non-ASD and high-risk ASD groups based on diagnostic outcome at 36 months.*

	Low-Risk		High-Risk		High-Risk Non-ASD		High-Risk ASD	
	M	SD	M	SD	M	SD	M	SD
7m Gross Motor	50.17	(8.98)	45.33	(9.98)	45.17	(8.83)	45.33	(12.66)
	n=50		**n=53**		**n=36**		**n=17**	
7m Visual Reception	54.73	(8.63)	50.73	(8.92)	51.71	(8.27)	50.87	(10.34)
	n=50		**n=53**		**n=36**		**n=17**	
7 m Fine Motor	57.79	(9.48)	52.49	(10.67)	53.54	(11.11)	49.67	(11.45)
	n=50		**n=53**		**n=36**		**n=17**	
7m Receptive Language	46.5	(8.73)	41.94	(11.71)	44.79	(10.38)	39.67	(12.20)
	n=50		**n=52**		**n=36**		**n=17**	
7m Expressive Language	50.29	(8.62)	43.2	(7.12)	41.67	(4.77)	45.2	(9.54)
	n=50		**n=53**		**n=36**		**n=17**	
14m Gross Motor	51.04	(16.25)	46.62	(16.16)	48.83	(20.52)	45.4	(17.59)
	n=50		**n=53**		**n=36**		**n=17**	
14m Visual Reception	55.85	(9.44)	52.16	(9.87)	52.35	(11.98)	49.47	(8.70)
	n=50		**n=53**		**n=36**		**n=17**	
14m Fine Motor	61.22	(9.19)	55.76	(11.89)	56.74	(12.44)	52.27	(11.95)
	n=50		**n=53**		**n=36**		**n=17**	
14m Receptive Language	46.36	(12.37)	43.94	(12.66)	48.04	(13.82)	42.27	(10.02)
	n=50		**n=53**		**n=36**		**n=17**	
14m Expressive Language	48.21	(9.41)	45.06	(11.32)	45.17	(11.24)	41.33	(11.56)

	n=50	**n=53**	**n=36**	**n=17**
24mGross Motor	59.89 (10.80)	45.19 (11.24)	44.59 (11.35)	44 (13.20)
	n=50	**n=51**	**n=35**	**n=16**
24m Visual Reception	58.93 (9.42)	53.59 (9.69)	56.21 (10.48)	52.93 (10.83)
	n=50	**n=53**	**n=36**	**n=17**
24m Fine Motor	54.33 (8.75)	50.06 (9.42)	51.21 (8.46)	47.33 (10.26)
	n=50	**n=53**	**n=36**	**n=17**
24m Receptive Language	59.14 (7.84)	51.1 (13.94)	51.21 (13.99)	46.8 (13.08)
	n=50	**n=53**	**n=36**	**n=17**
24m Expressive Language	58.16 (11.70)	49.94 (12.29)	49.17 (13.57)	47.2 (16.27)
	n=50	**n=53**	**n=36**	**n=17**
36m Visual Reception	59.03 (10.91)	56.42 (13.73)	60.17 (10.13)	50.64 (19.22)
	n=50	**n=53**	**n=36**	**n=17**
36m Fine Motor	56.58 (12.93)	49.45 (15.09)	54 (12.67)	41.20 (15.01)
	n=50	**n=53**	**n=36**	**n=17**
36m Receptive Language	57.4 (9.22)	50.9 (11.95)	54.13 (6.90)	46.87 (19.31)
	n=50	**n=53**	**n=36**	**n=17**
36m Expressive Language	58.98 (9.12)	53.16 (13.32)	55.42 (9.48)	47.87 (15.90)
	n=50	**n=53**	**n=36**	**n=17**

The Vineland Adaptive Behaviour Scales, Second Edition (VABS) was used as the second measurement of infant development level. This is a measure of adaptive behaviour in the form of interview questions and rating forms for the parent/caregiver, and can be used from birth to adulthood. The scales of the VABS are organised into communication, daily living, socialisation and motor domains. The communication domain assesses expressive, receptive and written communication skills. Daily living is used to evaluate the extent to which the child is integrated into the household, what they prefer to play with, and whether the child can eat and begin to dress themselves. The socialisation domain measures how a child interacts with other children and adults, how they cope in social situations, and how the child likes to play. The motor domain measures both gross and fine motor skills to manipulate objects and navigate within their environment. Table 2.2 provides the mean and standard deviation scores for low-risk and high-risk groups, and high-risk outcome groups at 36 months.

Table 2.2 *Mean and Standard deviation Vineland scores for high-risk and low-risk groups, high-risk non-ASD and high-risk ASD infants based on diagnostic outcomes at 36 months.*

	Low-risk		High-Risk siblings combined		High-Risk Non-ASD		High-Risk ASD	
	M	SD	M	SD	M	SD	M	SD
7m Communication	103.51	(13.93)	93.32	(15.38)	98	(17.08)	90.2	(11.9)
	n=50		**n=53**		**n=36**		**n=17**	
7m Daily Living	101.11	(15.56)	98.38	(16.46)	100.52	(18.95)	95.47	(15.90)
	n=50		**n=53**		**n=36**		**n=17**	
7m Socialisation	103.94	(12.43)	98.08	(14.91)	100.35	(15.23)	98.4	(15.36)
	n=50		**n=53**		**n=36**		**n=17**	
7m Motor	98.51	(14.51)	85.42	(16.06)	87.87	(17.27)	86.53	(17.18)
	n=50		**n=53**		**n=36**		**n=17**	
14m Communication	101.41	(11.22)	92.75	(16.71)	97.9	(12.57)	87.53	(15.46)
	n=50		**n=53**		**n=36**		**n=17**	
14m Daily Living	98.45	(9.84)	91.15	(12.93)	94.05	(10.04)	87.93	(13.17)
	n=50		**n=53**		**n=36**		**n=17**	
14m Socialisation	100.8	(9.88)	95.88	(14.37)	98.62	(11.59)	90.93	(14.95)
	n=50		**n=53**		**n=36**		**n=17**	
14m Motor	104.87	(12.14)	94.58	(13.63)	95.48	(10.23)	96.47	(15.28)
	n=50		**n=53**		**n=36**		**n=17**	
24m Communication	108.02	(12.97)	101.35	(12.74)	102.25	(13.63)	100.87	(14.96)
	n=50		**n=53**		**n=36**		**n=17**	
24m Daily Living	107.28	(10.93)	106.14	(10.36)	108.46	(9.18)	107	(9.71)
	n=50		**n=53**		**n=36**		**n=17**	
24m Socialisation	105.85	(12.09)	98.53	(9.19)	100.92	(8.48)	94.8	(9.34)
	n=50		**n=53**		**n=36**		**n=17**	
24m Motor	103.87	(9.51)	101.37	(10.31)	100.88	(9.62)	103.53	(12.06)
	n=50		**n=53**		**n=36**		**n=17**	
36m Communication	106.89	(10.94)	98.1	(14.26)	100.21	(11.79)	96.07	(19.31)
	n=50		**n=53**		**n=36**		**n=17**	
36m Daily Living	107.87	(8.31)	100.53	(12.26)	103.71	(9.12)	96.00	(17.97)
	n=50		**n=53**		**n=36**		**n=17**	
36m Socialisation	107.87	(9.15)	95.8	(13.72)	102.33	(9.22)	85.53	(13.53)
	n=50		**n=53**		**n=36**		**n=17**	
36m Motor	101.5	(8.31)	95.53	(11.98)	99.08	(11.19)	90.67	(14.29)
	n=50		**n=53**		**n=36**		**n=17**	

2.6.3 Autism outcome 36 months

Outcomes were determined using a comprehensive diagnostic evaluation. All infants were assessed using the ADOS-G *(ADOS-G:* Lord et al., 2000). The ADOS-G is a semi-structured assessment used in the diagnosis of ASD. The assessment focuses on play, social interaction and communicative domains. Semi-structured play sessions measure social behaviours with the aim of using social presses to produce social and communicative interactions and responses. Behaviours are coded between 0 and 3, and higher scores are indicative of a larger level of atypical behaviours that are associated with ASD. Cut-off points are then assigned overall, and individually in social and communicative domains. Individuals meet the criteria for ASD if they demonstrate scores above the cut off for ASD in both social and communicative domains. Parents of high-risk infants also completed both the Autism Diagnostic Interview Revised (ADI-R; Lord, Rutter, & LeCouteur, 1994), which was administered in the form of a semi-structured interview, and the Social Communication Questionnaire (SCQ; Rutter et al., 2003). In determining diagnostic outcome status, four experienced clinical researchers reviewed all information from 24-month (including ADOS-G assessments) and 36-month visits.

In this thesis, we combine typically developing individuals and those with 'other concerns' to form two outcome groups: high-risk ASD and high-risk non-ASD. Out of the 53 high-risk children assessed at 36 months, 17 (6 females, 11 males) met the criteria for an ASD diagnosis. A total of 36 children (26 females, 10 males) were classified as high-risk non-ASD. It is important to note here that the recurrence rate at 36 months in the BASIS sample was 32%. Whilst this is higher than the 19% reported in studies from the Baby Siblings Research Consortium (e.g. Messinger et al., 2015) other studies have reported consistently higher recurrence rates of 28% (Zwaigenbaum et al., 2016) and 29% (Elison et al., 2014) that

are in line with the BASIS results. It is also possible that the higher recurrence rate could, in part, reflect the modest sample size (Bedford et al., 2012; Shephard et al., 2016).

2.7 Results

First, between-subject (inter) variability was compared in the low-risk, and high-risk ASD outcome groups. In this case, the difference in variability was measured between groups using mean raw scores from Vineland and Mullen developmental measures.

Within-subject (intra) variability was measured in two ways. 1) A single score was created to represent variability across cognitive domains at a single time point for each individual, with the aim of identifying the extent to which an uneven cognitive profile was associated with an ASD outcome. This score was the standard deviation of all standardised cognitive measures at a single time point. A mixed ANOVA revealed the interaction of time-point and grouping on variability levels. Where significant group differences were identified, variability scores were correlated with Mullen and Vineland scores from 36 months to determine whether this type of variability was related to later performance levels. 2) Variability between time points was assessed in individual cognitive measures to determine whether the variability between two scores across time was associated with diagnostic outcome. One-way ANOVAs measured the relationship between variability (measured by taking the standard deviation between two time points) and outcome on each cognitive measure from the Mullen and Vineland cognitive assessments. Multiple regression models were used to examine the relationship between time-point variability and cognitive ability at 36 months, and whether this was related to group outcome.

2.7.1 Between-subject variability

Levene's Test for Equality of Variances was conducted to assess whether variability was higher in the high-risk ASD outcome group in comparison with individuals in low-risk and high-risk non-ASD categories using measures of motor, language, and communication domains. Variability in all Mullen (fine motor, visual reception, gross motor, expressive language and expressive language) and Vineland (socialisation, communication, daily living and motor) domains was compared between groups at 7, 14, 24 and 36 months. This resulted in 18 multiple comparisons, for which Bonferroni adjusted alpha levels of .0028 (.05/18) were used.

Levene's test revealed differences in variability between the high-risk ASD and low-risk groups in Mullen visual reception raw scores at 36 months (*SD* ASD = 7.0, *SD*. low-risk = 4.3, Levene's test $F = 22.1$, $p <$. 001), Mullen receptive language raw scores at 36 months (*SD* ASD = 9.5, Std. low-risk = 4.4, Levene's test $F = 14.8$, $p < .001$) and Mullen expressive language raw scores at 36 months (*SD* ASD = 8.5, *SD* low-risk = 3.8, Levene's test $F = 10.4$, $p < .001$). In all domains, heterogeneity was higher in the ASD group in comparison with low-risk outcome group. However, groups were not distinguishable at 7 months, 14 months, or 24 months. The earliest point that divergence of variability levels was demonstrated between groups was at 36 months (See appendix 1 for non-significant results of Levene's test).

Significant divergence in scores between outcome groups was identified only at 36 months across all cognitive domains. By 36 months, infants will have already been assessed and diagnosed with either ASD or other developmental issues such as delay. As variability scores were not significantly different at earlier time points (which may have provided an

opportunity to use variability as a predictive measure), further analyses were deemed unnecessary.

2.7.2 Within-subject variability

1) Measuring uneven cognitive profiles

In order to quantify heterogeneous development, a score was computed which represented the variability across domains within a single participant. Several studies have reported correlations between Mullen and Vineland measures (e.g. Luyster et al., 2008). Therefore, correlations were run between each of the behavioural parameters that were potentially to be included in the analysis. In total, Mullen fine motor, gross motor, visual reception, expressive language and receptive language domains were compared with Vineland communication, daily living, socialisation and motor scores. Table 2.3 shows all correlations above .3. A large correlation was considered to be a value of .5 or above (Cohen, 1988). Therefore, any variables at this level were considered for removal. Measures demonstrating the greatest variance were retained for analyses. Out of the correlated variables, Mullen scores demonstrated the greatest variance. Consequently, Vineland communication, motor, and daily living scores were removed from further analyses. Mullen gross motor, visual reception, fine motor, expressive language, receptive language and Vineland socialisation scores were retained.

Table 2.3 Correlated Mullen and Vineland subdomain variable using 36-month data. M= Mullen, V= Vineland. All correlations were significant p <. 005.

	M Gross Motor	M Fine Motor	M Visual Reception	M Receptive Language	M Expressive Language
V Communication	.309	.330	.345	.643	.596
V Daily Living	.458			.379	.419
V Socialisation					.349
V Motor	.739	.353		.333	.355

Mixed ANOVAs

Z-transformations were conducted on all data in order to standardise scores from different testing scoring scales. Scores from the five sub-domains of the Mullen, and Vineland socialisation were transformed across the 4 time-points (7, 14, 24 and 36 months), resulting in 6 domain values per time point, per individual. In order to create a within-subject variability score, the standard deviation of these six domains was computed for each time-point. This resulted in one value per time point per participant that represented variation across the cognitive profile.

The computed within-subject variability score was analysed in a mixed four within (time points) by three between (groups: low-risk, high-risk non-ASD and high-risk ASD) ANOVA to identify statistical differences in within-subject variability scores between the three outcome groups across development. There was homogeneity of covariance, as assessed by Box's test (p = .671). The data violated the assumptions of normality for ANOVA, as measured by the Shapiro-Wilk's test. However, ANOVA is reasonably robust to deviations from normality. Furthermore, there is no parametric alternative that allows the inspection of interaction effects. In this case, I was interested in the interactions between time point and outcome group, and the levels of variability. Therefore, the analysis was run irrespective of this violation.

The ANOVA revealed a significant interaction effect between time point and grouping on between-domain variability levels $F(9,264) = 26.707$, $p = .006$), $n_p^2 = .182$. There was a significant overall main effect of time point $F(3,89) = 3.602$, $p = .001$), $n_p^2 = .109$. There was also a significant main effect of outcome group (3,75) = 57.707, $p = .001$), $n_p^2 = .416$. Simple main effects are shown below.

Effect of time point

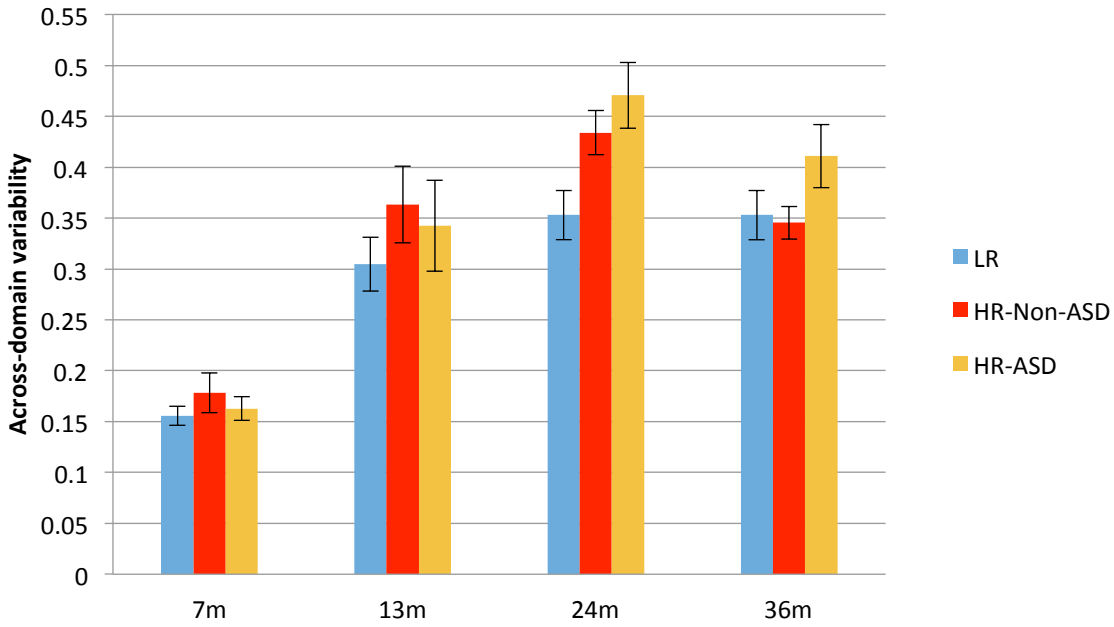
There was a significant effect of time on variability scores between domains for the low-risk group $F(3,123) = 19.883$, $p = .001$), $n_p^2 = .327$. The low-risk group demonstrated higher between-domain variability levels at 14 months (M = .151, SD = .028, $p = .001$) in comparison with 7 months but remained stable after this time. There was a significant effect of time for the high-risk non-ASD group $F(3,60) = 17.597$, $p = <.001$), $n_p^2 = .468$. The high-risk non-ASD group demonstrated significantly higher variability at 14 months (M =.187, SD =.029, $p = .001$) in comparison with 7 months, and lower levels at 36 months (M = -.144, SD = .025, $p = <.001$) compared with 24 month scores. The high-risk ASD showed a significant effect of time point $F(3,42) = 21.507$, $p = <.001$), $n_p^2 = .685$. Variability between domains was significantly higher at 14 months (M = .196, SD =.042, $p = .002$) compared to 7 months, and was also significantly higher at 24 months (M = .134, SD = .037, $p = .015$).

Effect of group

There was statistically significant difference in variability between outcome groups at 24 months $F(3,91) = 5.089$, $p = .003$), $n_p^2 = .271$. Variability was significantly larger in the high-risk ASD group (M = 2.066, SD = .054, $p = .001$) compared to the low-risk group at this

point. However, no significant differences were identified between high-risk ASD and high-risk Non-ASD groups (M = .130, SD = .035, p = .067). Figure 2.1 shows the mean scores and standard errors across the four time points for each outcome group.

Figure 2.1. Mean variability scores across cognitive domains in low-risk, high-risk non-ASD and high-risk ASD populations. Error bars represent the standard error of the mean.



Correlations

Correlations were run on the whole sample in order to assess the relationship between within-participant variability at 24 months and developmental measures at 36 months. These measures were the four Mullen outcome variables (raw, note that Mullen gross motor was not conducted at 36 month assessments) and the four Vineland scores (raw) of time point 4 (resulting in 10 multiple comparisons, for which p values were Bonferroni corrected to a significance value of .005.

One outlier was present in the overall correlations as assessed by inspection of a boxplot. The value was greater than 2.5 lengths from the edge of the box and defined as an extreme outlier. It was therefore removed rather than modified. An increased within-subject variability score overall at 24 months was correlated with negative developmental outcomes at 36 months across social and communicative domains (after correcting for multiple comparisons): Mullen visual reception (r = -.298, n = 100), Mullen receptive language (r = -.272, n = 98, $p < .01$), Mullen expressive language (r = -.321, n = 97, $p < .01$) and Vineland Socialisation (r = -.338, n = 101, $p < .01$)

Correlations were then conducted across high-risk and low-risk groups, and across all high-risk subgroups in order to see whether adverse outcomes were associated with ASD at 36 months, or whether high-risk subgroups were indistinguishable. The only significant correlations were in the high-risk group when not grouped by outcome. The correlations showed an association between higher domain variability at 24 months and lower scores in Mullen visual reception (r = -.393, n = 51, $p < .01$), Mullen expressive language (r = -.379, n = 51, $p < .01$), and Mullen receptive language (r = -.349, n = 51, $p < .01$) domains at 36 months. See table 2.4 for all results.

Table 2.4 Correlations between variability scores in low-risk and high-risk groups and 36-month cognitive ability scores. M = Mullen, V = Vineland raw scores.

		Overall	Low-risk	High-Risk	High-risk non-ASD	High-risk ASD
M Visual reception	Cor. coefficient	-.291	-.047	-.393	-.092	-.425
	p value	**.002***	ns.	**.001***	ns.	ns.
	n	99	49	51	36	15
M Expressive language	Cor. coefficient	-.321	-.061	-.379	-.193	-.352

	p value	**.001***	ns.	**.001***	ns.	ns.
	n	99	49	51	36	15
M Receptive Language	Cor. coefficient	-.290	-.036	-.349	-.013	-.295
	p value	**.002***	ns.	**.001***	ns.	ns.
	n	99	49	511	36	15
V Socialisation	Cor. coefficient	-.292	-.95	-.195	-.154	-.069
	p value	**0.001***	ns.	ns.	ns.	ns.
	n	99	49	51	36	15

The correlations suggest a divergence between high-risk and low-risk infants at 24 months. This was identified as an overall increase in variability across cognitive domains, which was indicative of an uneven cognitive profile. A higher variability score was associated with lower social and communicative abilities at 36 months as assessed by Mullen and Vineland scores. Interestingly, variability levels in individuals with a diagnosis of high-risk non-ASD and high-risk ASD were not significantly different. One possibility for the elevated levels of variability shown in the high-risk non-ASD category is that these individuals could be exhibiting some sub-clinical characteristics of ASD, but not at the level of individuals with the disorder. This could explain why both groups exhibited higher levels of variability than low-risk infants.

A logistic regression was performed to ascertain the effect of variability on 36-month outcome scores. However, the logistic model was non-significant X^2 (1) = .809, p = .289.

2.7.3 Within-subject variability between time points

Within-subject variability was also analysed by comparing changes between the four time points at which infants were assessed. Variability in timing was assessed in all available cognitive measures. Standardised Z-scores from Mullen and Vineland measures of infant

development were utilised in order to measure ability levels over a range of cognitive measures domains. Scores from 7, 14, 24 and 36 months were included in the analyses. In order to compare the change in individual variability over time, three time point contrasts were created; the change in variability between 7 and 14 months, 14 and 24 months, and 24 and 36 months. The standard deviation was used as the measurement of variability between the two time points. Each score was represented by the change in variability for one measurement. One-way ANOVAs were conducted to determine whether between-time variability was distinct in the three outcome groups, low-risk (n = 50), high-risk non-ASD (n = 36) and high-risk ASD (n = 17). Significant differences were identified in two cognitive measures. Significant results are presented below.

Expressive language

Variability between 7 and 14 months

There was no statistically significant difference in variability score between the outcome groupings, $F(2,96) = .325$, $p = .724$.

Variability between 14 and 24 months

A one-way ANOVA was also conducted looking at variability between visual reception scores between 14 and 24 months. There were no outliers, as assessed by the boxplot test. The variability score was not normally distributed when assessed by Shapiro-Wilk's test ($p > .05$). However, as non-normality does not considerably affect Type I error rates and one-way ANOVA is reasonably robust to non-normality (Maxwell & Delaney, 2004), the analysis was run without alteration. There was homogeneity of variances, which was assessed by Levene's test for equality of variances ($p = .130$).

Variability between 14 and 24 months increased from the low-risk (n = 48, M =. 373, SD = .309) to high-risk non-ASD (n = 36, M = .665, SD = .414) to high-risk ASD (n = 15, M = 1.686, SD = .576) outcome groups. The difference in within-subject variability scores was significantly different for levels of outcome group, $F(2,96) = 71.553$, $p < .001$. Tukey-Kramer was used for post hoc analyses, as case numbers were unequal in the outcome groups. The increase in variability between 14 and 24 months from low-risk to high-risk non-ASD (M = .293, SE = .083) was significant ($p < .05$). The increase in variability between high-risk non-ASD and high-risk ASD (M = 1.021, SE = .114) was significant ($p < .05$). The increase in variability between low-risk and high-risk ASD (M = 1.314, SE = .109) was also significant ($p < .05$).

Variability between 24 and 36 months

There were no outliers, as assessed by the boxplot test. Variability score was not normally distributed when assessed by Shapiro-Wilk's test ($p > .05$). However, as in the 14 to 24 month analysis, the ANOVA was run without alteration. There was homogeneity of variances, which was assessed by Levene's test for equality of variances ($p = .210$).

Variability between 24 and 36 months increased from the low-risk (n = 48, M =. 109, SD = .091) to high-risk non-ASD (n = 36, M = .421, SD = .341) to high-risk ASD (n = 15, M = .864 SD = .381) outcome groups. The difference in within subject variability scores was significantly different for levels of outcome group, $F(2,96) = 42.595$, $p < .001$. Tukey-Kramer post hoc analysis showed that the variability between 24 and 36 months from low-risk to high-risk non-ASD (M = .311, SE = .064) was significant ($p < .05$). The increase in variability between high-risk non-ASD and high-risk ASD (M = .443, SE= .062) was

significant ($p < .05$). The increase in variability between low-risk and high-risk ASD (M = .755, SE = .084) was also significant ($p < .05$). Figure 2.2 shows the significant time points.

Figure 2.2 Expressive language variability scores for each outcome group between 14-24 and 24-36 months. Error bars represent the standard error of the mean.

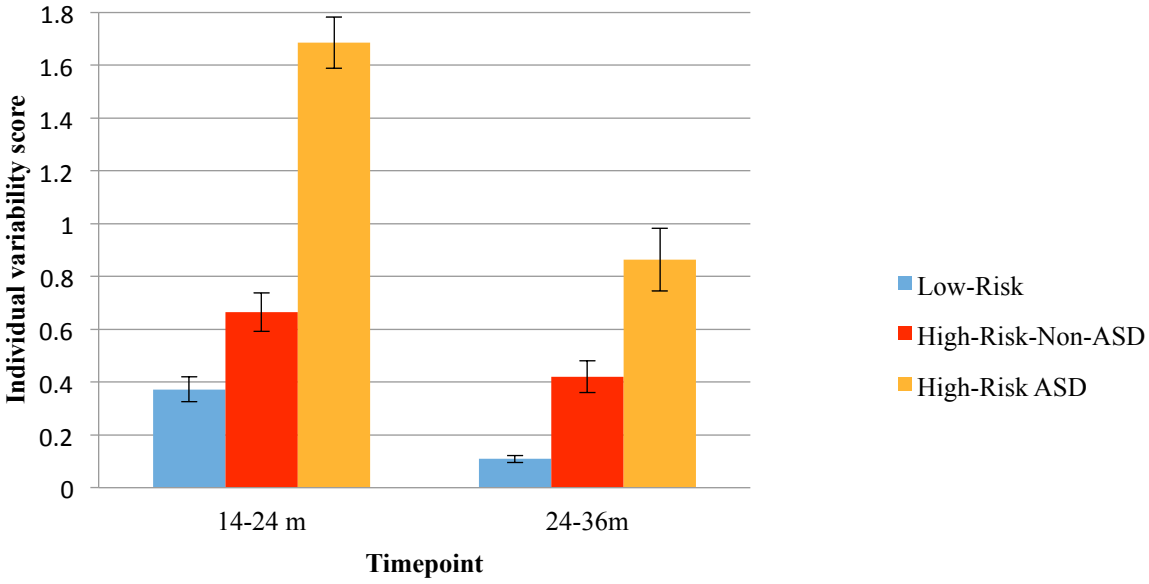


Figure 2.2 shows that between 14 and 24 months, the variability in expressive language scores was significantly higher in the high-risk populations compared to the low-risk outcome group. Furthermore, the high-risk ASD group exhibited a larger variability score compared to individuals in the high-risk-non-ASD category.

Visual reception

Variability between 7 and 14 months

The variability between 7 and 14 months decreased from the high-risk ASD (n= 15, M = .447, SD = .308) to high-risk non-ASD (n = 36, M = .649, SD = .397) to low-risk (n = 48, M = .680, SD = .516) outcome groups. However, there was no statistically significant difference in variability score between the outcome groupings, $F(2,96) = 01.457$, $p = .238$.

Variability between 14 and 24 months

There were no outliers, as assessed by the boxplot test. The variability score was normally distributed as assessed by Shapiro-Wilk's test (p = >.05). There was homogeneity of variances, which was assessed by Levene's test for equality of variances (p = .132). Variability between 14 and 24 months increased from the high-risk ASD (n = 15, M =. 297, SD = .223) to low-risk (n = 48, M = .596, SD = .556) to high-risk non-ASD (n = 36, M = .667 SD = .347) outcome groups. The difference in within subject variability scores was significantly different for levels of outcome group, $F(2,96) = 3.264$, $p = .043$.

Tukey-Kramer post hoc analyses showed that the decrease in variability between 24 and 36 months from high-risk ASD to high-risk non-ASD (M = .370, SE = .146) was significant (p <.05), but no other group differences were significant.

Variability between 24 and 36 months

There were no outliers, as assessed by the boxplot test. The variability score was normally distributed as assessed by Shapiro-Wilk's test (p = >.05). There was homogeneity of variances, which was assessed by Levene's test for equality of variances (p = .091). Variability between 24 and 36 months decreased from the high-risk ASD (n = 15, M =. 169, SD = .154) to high-risk non-ASD (n = 36, M = .350, SD = .223) to high-risk non-ASD (n =

48, M = .407, SD = .339) outcome groups. The difference in within subject variability scores was significantly different between levels of outcome group, $F(2,96) = 3.636$, $p = .031$.

Tukey-Kramer post hoc analyses showed that the decrease in variability between 24 and 36 months from high-risk ASD to low-risk groups (M = .238, SE = .088) was significant ($p = .023$), but no other group differences were significant. Figure 2.3 shows the significant mean differences between outcome groups.

Figure 2.3 Visual reception variability scores for each outcome group between 14-24 and 24-36 months. Error bars represent the standard error of the mean.

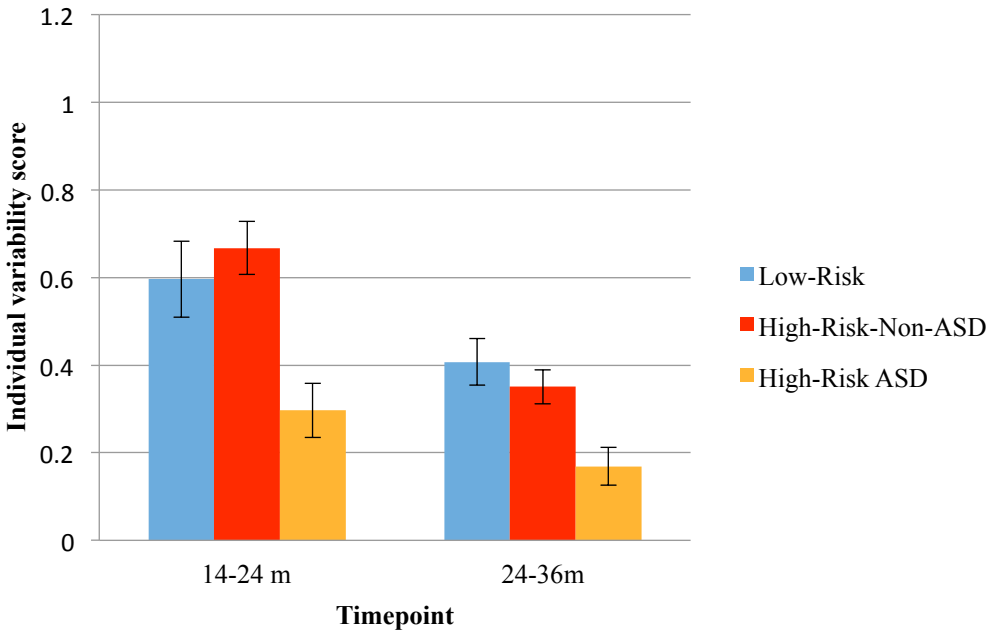


Figure 2.3 shows that between 14 and 24 months, the variability in visual reception scores was significantly lower in the high-risk populations compared to the low-risk outcome group. Furthermore, the high-risk ASD group showed a significantly lower variability score

compared to individuals in the high-risk non-ASD category. The same pattern was demonstrated at 36 months, but by this point all groups showed reduced variability.

Variability as a predictor of development.

Individual variability scores between 14 and 24 months were significantly different across outcome groups in expressive language and visual reception domains. The high-risk ASD group displayed a significantly higher variability score in expressive language compared to low-risk and high-risk non-ASD groups. The high-risk group also displayed a significantly lower variability in visual reception in comparison with the LR and HR-non ASD groups. Multiple regressions were run to examine whether variability score in expressive language was indicative of expressive language scores at 36 months based on outcome groupings, and whether variability in visual reception was associated with visual reception scores at 36 months. However, both models were non-significant. Visual reception $F(1, 98) = 3.393$, $p = .122$, adj. $R^2 = .26$, expressive language $F(1, 98) = 2.45$, $p = .265$, adj. $R^2 = .34$.

2.8 Discussion

The aim of this chapter was to investigate within- and between-subject variability utilising data from a high-risk, prospective dataset. Infants in low-risk, high-risk non-ASD and high-risk ASD outcome groups were compared across cognitive measures between 7 and 36 months to examine the group relationship between variability and cognitive abilities over time. Within-subject variability was then examined in two ways. First, variability between cognitive domains was measured and compared at 7, 14, 24 and 36 months in each individual to elucidate the extent to which groups could be differentiated in terms of an uneven cognitive profile. A second within-subject variability measure compared variability between time points in language, motor and communicative domains. Between-group variability was

not predictive of outcome. However, evidence from analyses of within-group variability demonstrated initial evidence of predictive power. I discuss implications from current results and future work below.

Levene's test for equality of variances was used to compare variability levels between the outcome groups across social, communicative and motor domains. The primary aim was to identify the earliest time point at which variability significantly differed between outcome groups. The high-risk ASD group displayed higher levels of variability in visual reception, expressive language and receptive language domains compared to both low-risk and high-risk non-ASD groups. However, this effect was present only at the final testing point, 36 months, so could not be used to predict later developmental outcomes. This result is not surprising given the highly heterogeneous nature of ASD. However, due to the fact that atypicalities are identified in high-risk studies by 12 months across many cognitive and behavioural domains (see Chapter 3 for a literature review of high-risk prospective studies), it was possible that variability may have been identified earlier.

Heterogeneity within the ASD group increased over time, as assessed by the standard deviation across cognitive measures between 7,14, 24 and 36 months. This, combined with significantly higher variability levels at 36 months was concordant with findings from previous prospective research (Landa et al., 2013; Charman et al., 2005) that reported increasing heterogeneity levels across multiple cognitive domains in ASD groups compared to low-risk groups. One explanation for the increased levels of heterogeneity is that ASD could encompass multiple underlying disorders. Whilst problems with communication, social interaction, and restricted patterns of behaviour or interests with varying levels of severity define all individuals with ASD, ASD can have multiple diagnoses. The DSM-5

(American Psychiatric Association, 2013) now comprises three severity levels in the diagnostic criteria, and includes individuals with the once distinct Asperger's disorder. Moreover, as previously discussed, an increasing number of researchers have attempted to identify homogenous subgroups within ASD. Some research has proposed that distinct behavioural subtypes based on behavioural profiles would be beneficial to identify subgroups on a wide spectrum (for example, Ozonoff et al., 2010; Landa & Garrett-Mayer, 2006; Landa, Gross, Stuart & Bauman, 2012). Others argue that we must move away from conceptualising ASD as a unitary disorder, instead viewing ASD as a syndrome consisting of multiple distinct disorders, where multiple causal pathways can be identified (Ronald et al., 2006; Geschwind & Levitt, 2007). Robust evidence has not yet been provided for either account of ASD. However, if ASD were to be identified as a disorder that could encompass multiple diagnoses, then it would, by definition, include more variability.

When intra-variability across domains was examined, a significant effect of time and group were identified. The low-risk group displayed a significant increase in variability between 7 and 14 months. However, between 14 and 36 months, variability stabilised. Between 7 and 24 months, both individuals in the high-risk non-ASD and high-risk ASD groups exhibited increases in heterogeneity. Whereas by 36 months, only high-risk ASD variability levels significantly decreased. These patterns of variability suggest that regardless of group, infants demonstrated similar levels of heterogeneity at 7 months. This finding is congruent with reports from high-risk prospective studies that individuals with ASD are indistinguishable from typically developing children at 7 months in social (Young et al., 2009; Rozga et al., 2009), communicative (Cassel et al., 2007) and language (Lazenby et al., 2015; Landa et al., 2007) domains. If comparative levels of variability and performance are seen in the ASD and typically developing groups, it demonstrates that a more uneven cognitive profile (indicating

strengths in some cognitive domains and difficulties in others) has not yet developed, and therefore suggests that cognitive and behavioural abilities may still be within the typical range. However, by 14 months, all high-risk populations in the high-risk sample exhibited more uneven cognitive profiles, suggesting divergence from a typical developmental trajectory. The Over-Pruning account offers an explanation for a period of (seemingly) typical development in early infancy. The account proposes that over-aggressive synaptic pruning is the pathological mechanism in ASD. As synaptic pruning does not commence before 6 months of age, the account predicts that there should be a phase of typical development prior to the emergence of atypical behaviours. This pattern of development is reported in the current chapter.

The earliest significant difference in heterogeneity levels between the low-risk and high-risk groups was at 24 months. At this point, individual variability was larger in the high-risk group overall, but not between high-risk non-ASD and high-risk ASD groups (see figure 2.1). Furthermore, when correlations were conducted between groups to assess the association of heterogeneity at 24 months and cognitive ability at 36 months, negative correlations were reported in the high-risk group overall, but not high-risk ASD. In the high-risk group, higher levels of heterogeneity across domains at 24 months predicted negative developmental outcomes at 36 months in visual reception and language domains. Taken together, significant differences were not identified between high-risk groups. It is possible that this is evidence of the broader autism phenotype, proposed by Ozonoff et al., (2010) (see Chapter 3). In this case, individuals with ASD demonstrate sub-clinical symptoms of ASD in some cognitive domains. The BAP is also more likely to be identified in individuals from a high-risk family (Ozonoff et al., 2014). Therefore, an uneven cognitive profile could also be present in high-risk individuals who are not diagnosed with ASD. This result is consistent with hypotheses

from the Over-Pruning account of ASD. The account proposes that two separate factors will impact the development of unaffected high-risk siblings, leading to differences from typically developing controls: either through inheriting a milder version of the pathological cause of ASD (in this theory, over-aggressive synaptic pruning) or through inherited risk factors that would cause a different trajectory of development. Both cases could cause a sub-clinical phenotype.

It would be of interest to ascertain whether within-subject variability at 24 months could be a predictor of later development. Anderson et al. (2008) measured language abilities at four time points between 2 and 9 years in children with ASD. Variability was high at 2 years and increased over time in the ASD group. However, symptom severity and non-verbal abilities at the age of 2 were predictive of expressive and receptive language outcomes at 9 years. Here, the correlations showed an association between high variability across domains and lower ability scores in visual reception and language domains. Thus, it would be of interest to apply this type of analysis to data with later outcomes.

A binary logistic regression analysis was run to establish whether variability across domains at 24 months was predictive of diagnostic outcome at 36 months. However, the regression was non-significant. Whilst the variability measure was not sufficient in predicting diagnostic outcomes when used as an independent measure, one future direction would be to utilise it in conjunction with other predictors. An example of this would be to utilise variability scores in cluster analysis aiming to identify homogenous subtypes within ASD. Kim, Macari, Koller and Charwarska (2016) argue that a variety of measures should be incorporated into models, rather than measures that only assess the severity of ASD

symptoms. A future direction would be to evaluate the predictive value of this form of intra-subject variability when combined with other measures.

ANOVAs comparing variability between 14 and 24 months demonstrated group differences in expressive language and visual reception domains in the high-risk ASD group. As a group, high-risk ASD infants demonstrated a larger variability between time points. Conversely, between 14 and 24 months the high-risk ASD group showed significantly reduced variability in visual reception. It must be questioned why variability might be higher in some cognitive domains compared to others. Looking to evidence in the literature that could support these findings, a putative discrepancy between visual and language processing abilities has been reported (Behrmann, Thomas & Humphreys, 2006). Specifically, a number of studies have reported less atypicality in visual processing abilities in individuals with ASD (Caron, Mottron, Berthiaume, & Dawson, 2006; de Jonge et al., 2007). More recently, Sahyoun et al. (2009, 2010) used functional magnetic resonance imaging (fMRI) in conjunction with a picture-based problem-solving task whilst manipulating the availability of language-based and visual information to solve the problems. The authors reported an increased reliance on visual processing strategies and a decreased engagement of language mediation in children with ASD. Therefore, if infants exhibit higher levels of variability in expressive language tasks (due to an increased level of atypicality within this domain), this could explain why variability could be lower in visual perception. However, this study does not utilise young children with ASD. Therefore, at this point, we cannot elucidate the developmental processes that led to this cognitive profile, and when these patterns of behaviour were first present.

Alternatively, a decrease in variability levels could indicate lower ability scores in ASD. It is possible that variability between time points is lower (and therefore performance levels are more stable at a group level) because individuals with ASD are demonstrating floor effects in visual perception tasks. However, as a group, individuals with ASD have been shown to exhibit atypicalities in expressive language between 14 and 24 months, so it seems questionable that floor effects could be the cause here. In order to assess whether variability between time points is due to a higher or lower ability level, it will be important to consider individual scores in the Mullen visual reception assessment and compare scores between time points.

Regression models were then run to ascertain whether variability between 14 and 24 moths was associated with abilities in expressive language and visual reception domains at 36 months. However, no associations were identified were later outcomes.

One limitation was the relatively small sample size, particularly in the high-risk ASD group. It is possible that lack of significance when the high-risk groups were split was due to the reduced power from the smaller sample size. One way to overcome this problem would be to combine data from a second cohort of infants recruited in BASIS known as Phase 2. At the time of testing, Phase 2 data was unavailable. However, for future work, this could be one method of increasing statistical power.

A second limitation was that measuring between-group heterogeneity caused some significant differences to be masked by mean effects. Comparing standard deviation scores from the low-risk group against the high-risk groups on single measures can be misleading. In general, individuals with ASD display larger levels of variability in individual scores, and

this variability could be sufficient to eliminate significant differences between groups. Future work should therefore combine analyses at the group and individual level in order to ensure that group variability scores are representative of the whole sample. A final point is whether Mullen and Vineland were the 'best' measures to show variability. It is possible that test measures that contain a narrower range of assessments with the capability to capture more fine-grained behaviours in particular domains could identify variability where developmental assessments such as the Mullen and Vineland have not.

A third limitation is the use of standard deviation as a measure of variability. Whilst a proportion of the estimated variance can be accounted for by intra-subject variability, it must be questioned whether difficulties in assessing young children with ASD could also lead to higher levels of variability in the sample. Assessing young children who exhibit features of ASD can be particularly challenging, and a number of behaviours can hinder accurate assessments. Many children often display more limited social interaction and communication abilities (Marcus, Lansing & Schopler, 1993). Other problems include sensory issues, over activity and poor compliance (Akshoomoff, 2006). Therefore, it is possible that such difficulties might lead to higher levels of variability. One way to assess the extent of measurement error would be to compare results with parental reports and semi-structured interviews.

This chapter provides some of the first analyses to utilise variability scores between groups, and within individuals over time. Variability scores were calculated to measure inter-measure variability between 7 and 36 months. It would be valuable to measure inter-variability from early to mid-development (for example, between 3 years and 7 years) to determine whether it could provide predictive power later in development. Turning to intra-

subject variability, a higher variability score (more uneven cognitive profile) at 24 months was correlated with negative outcomes in language and communicative domains at 36 months in the high-risk group overall, but was not a predictor of outcome. Understanding the differences in variability levels between high-risk infants who are diagnosed ASD and high-risk siblings who are not diagnosed could benefit our understanding of the broader autism phenotype. It would, for example be interesting to look at the development of cognitive profiles later through early childhood in order to identify the point at which high-risk non-ASD individuals may demonstrate less uneven cognitive profiles than individuals with ASD. ANOVAs comparing measures of variability between time points in single cognitive domains demonstrated differences in expressive language and visual reception domains. A pattern of higher variability in expressive language and lower variability in visual perception domains were identified the high-risk ASD. This led to the question of whether infants were exhibiting higher ability levels in visual processing, or whether infants were exhibiting floor effects in visual reception tasks.

Despite several limitations, these findings provide evidence of the utility of variability measures in further understanding the development of ASD. It demonstrates the importance of elucidating the neurobiological processes underlying the disorder. For example, understanding whether ASD is comprised of one or several disorders with distinct causal mechanisms will greatly benefit the way that we understand and account for variability. Furthermore, utilising both inter-subject and intra-subject variability in high-risk infants, alongside the analysis of cognitive and behavioural measures across development could provide insight into the divergence in individuals who go on to develop ASD and those who exhibit features consistent with the broader autism phenotype.

Chapter 3

3.1 Introduction

As discussed in chapter 2, heterogeneity or variability is present in both the manifestation of symptoms and aetiology in ASD. By three years of age, individuals with ASD exhibit difficulties across multiple domains, and it is likely that a range of atypicalities may be present in early infancy. The current chapter focuses on the identification of atypical behaviours between 7 and 14 months in at-risk infants. Regression models incorporated data from multiple cognitive and environmental domains from the BASIS prospective longitudinal study of ASD, as predictors of developmental outcome at 24 and 36 months. In order to validate significant predictive results, regression models were replicated with data from a second, independent cohort from the BASIS network. Here, I review prospective findings identifying early cognitive and behavioural atypicalities in high-risk infants that have been associated with a later ASD diagnosis. The current findings are then discussed in relation to predictions from the Over-Pruning account of ASD.

3.2 Predictors of later outcome in ASD in prospective research

The development of ASD across the first years of infancy has been identified as a crucial area of research in recent years (Meng-Chuan et al., 2016). Understanding the developmental patterns of emerging symptoms provides the potential to improve the implementation of intervention strategies and also enables researchers to further understand the complex interplay of developmental processes underlying ASD. Thus far, longitudinal prospective studies have provided the most convincing evidence of behavioural and cognitive markers in the first 18 months of life that could predict ASD outcomes. Therefore, the following review focuses on prospective findings of early predictors.

3.2.1 Early social predictors of outcome

Impairments in social interactions are a primary diagnostic feature of ASD. This includes difficulties with reciprocal interactions and the appropriate use of gesture and facial expressions, and a repetitive repertoire of interests and activities (DSM-5, American Psychiatric Association, 2013). Many prominent causal theories of autism are centred around the proposal that the earliest impairments are exhibited in social domains. It is argued that secondary atypicalities in the form of sensory and motor problems are caused by a reliance on the typical development of primary social behaviours, which have, in the case of individuals with ASD, developed atypically (Dawson, Meltzoff, Osterling & Rinaldi, 1998; Chevallier, Kohls, Troiani, Brodkin & Schultz, 2012). Such theories argue for a particular order of the manifestation of ASD symptoms, beginning with social domains. However, the consensus from prospective studies thus far is that infants who are later diagnosed with ASD are indistinguishable at 6 months in terms of social interactions when compared with typically developing infants (Young et al., 2009; Rozga et al., 2011; Bedford et al., 2012). The 'still face' paradigm is one experimental procedure that has been used in prospective studies to measure social capabilities of high-risk infants. Here, interactions from a caregiver are withdrawn from an infant, at which time the infant's sensitivity to the changing situation is measured. Young et al. (2009) found that gaze behaviours were typical at 6 months in all infants who were later diagnosed with ASD. Furthermore, infants who at 6 months exhibited atypical gaze patterns (that is, looking less to the eye region during the still face paradigm task), did not go on to develop ASD or developmental problems and were categorised as typically developing at 36-months.

Using the still face paradigm Rozga et al. (2011) reported no differences between the ASD

group, high-risk non-ASD group, or low-risk group at 6 months in frequency of gaze, smiles or sensitivity as a result of withdrawal of parent interaction, with all groups demonstrating a typical reduction in gaze towards the caregiver. Thus, increased attention to the face and decreased attention to the eyes was not a stable characteristic in 6-month old infants later diagnosed with ASD. However, the authors argue that at 12 months, infants who were later diagnosed with ASD could be differentiated from low-risk and high-risk non-ASD infants on the basis of disturbances in socially directed behaviours. Specifically, pointing, showing and requesting behaviours were delayed in the later diagnosed ASD group.

Response to interactions in the form of social attention and facial processing has also been explored in prospective research. Research focusing on joint attention (a social-communicative skill where an infant and caregiver employ gaze and gesture behaviours in order to interact with object in an environment) has identified differences in high-risk infants that go on to develop ASD, but these atypicalities have been found no earlier than in other non-social domains. Walden, Stone and Yoder (2007) assessed at-risk infants between the ages of 12 and 23 months on a range of joint attention prompts involving gaze shifting and pointing. They found that initial levels of joint attention at 15 months predicted later impairments in joint attention and ASD diagnosis. Furthermore, Cassel et al. (2007) assessed clinical observations of social and communicative behaviours in high-risk infants using the Early Social Communication Scale. Cassel and colleagues did not identify impairments in joint attention when infants were evaluated at 8,10,12, 15 months. Impairments were only identified later in development, at 18 months.

Bedford et al. (2012) suggest that joint attention behaviours emerge over the second year of life and are not present from birth (see also Butterworth & Jarrett, 1991; Tomasello,

Carpenter, Call, Behne & Moll, 2005; Mandy, Sullivan & Mastergeorge, 2009). Therefore, understanding precursor behaviours to joint attention such as gaze following may provide information about the underlying mechanisms of social and communicative impairments. Eye tracking was used to explore the differences in gaze following and looking behaviours between high-risk and low-risk infants at 7 and 13 months. Infants viewed videos of an actor turning to look at one of two objects. Gazed-at objects were known as "congruent", whilst non gazed-at objects were known as "incongruent". Bedford et al. (2012) reported that differences between groups in the proportion of correct looks (to the congruent object) were not present at 7 or 13 months. The authors suggest that mechanisms for automatic orienting to another's gaze are therefore intact early in development. In the second part of the task, the actor in the videos engaged the watching infant in eye contact before shifting their gaze to an object. In typically developing infants, the number of looks towards the congruent object was significantly higher than that to the incongruent object after watching the model making eye contact. By 13 months however, infants later diagnosed with ASD or those who showed some level of social or communicative impairment but did not meet criteria for ASD, spent significantly less time viewing the congruent object after gaze following. The interpretation from the authors is that after correctly following gaze direction, infants who exhibited social-communication difficulties later in development may not be using information from attentional engagement to attend preferentially to the gazed-at object in the task. This reduction in attention could reflect difficulties in understanding the importance of eye-gaze in communicative tasks, which could in turn contribute to the on-going developmental process that leads to ASD symptomology.

Research focusing on attention to unfamiliar faces has proved contradictory in regards to the time at which atypicalities in this area are demonstrated in infants later diagnosed with ASD.

For example, no group differences were identified in 6-month-old infants in facial gaze or smiling and vocalisation behaviours when tested in a naturalistic setting (Ozonoff et al., 2010), and no differences were identified at 6 or 12 months in visual attention towards unfamiliar faces in naturalistic settings using an eye-tracking task (Elsabbagh et al., 2013a). Conversely, Chawarska et al. (2013) reported that infants who were later diagnosed with ASD spent less time attending to adult faces at 6 months. The authors assessed social monitoring skills in high-risk and low-risk infants using eye-tracking techniques. Videos of an actor engaged in several activities that varied in social content (for example, speaking to the camera or making a sandwich). Compared to the control group, infants who were later diagnosed with ASD exhibited a lower ability to attend spontaneously to people and spent significantly less time monitoring the actor, and in particular, the face area.

Similarly, Nele, Ellen, Petra and Herbert (2015) showed that whilst there was no difference between high-risk and low-risk infants displaying a preference for their mother's face at 5 months, only low-risk infants showed a preference for faces with direct gaze over faces with averted gaze; high-risk infants did not discriminate between the categories. Jones et al. (2014) argue that in order to further explain discrepancies in prospective findings, it is vital to understand the origins of impairments, and how early communicative atypicalities may constrain learning in other social and communicative domains.

3.2.2 Language

Delays in both language comprehension and language production have been identified in early infancy. A delay in the production of first words has been identified as one of the primary markers of atypicality in infants who are later diagnosed with ASD (Mitchell, 2006; Landa & Garrett-Mayer, 2006), and numerous high-risk sibling studies have identified delays

in high-risk infants at 12 months that were indicative of an ASD diagnosis at 24 or 36 months. Zwaigenbaum et al. (2005) analysed scores from the Mullen Scales of Early Learning (MSEL), comparing high-risk and low-risk infants. The MSEL is an assessment battery designed to measure infant development. It contains two language subscales and measures both receptive and expressive language skills. Subgroups defined at 24 months using ADOS classifications displayed differences in expressive language scores at 12 months. Siblings classified with ASD demonstrated lower scores than both low-risk infants and typically developing high-risk siblings. Furthermore, results from the MacArthur Communicative Development Inventories-Words and Gestures (CDI) parental questionnaire showed that infants later diagnosed with ASD also exhibited fewer understood phrases at 12 months in comparison with other groups. Similarly, Landa and Garrett-Mayer (2006) reported lower MSEL expressive language scores in 14-month infants later diagnosed with ASD at 24 months (see also Landa, Holman & Garrett-Mayer, 2007).

Delays have also been identified in language comprehension. Zwaigenbaum et al. (2005) identified differences at 14 months on the CDI parental report in language comprehension. Infants diagnosed with ASD at 24 months understood fewer phrases and gestures between 12 and 14 months. Moreover, Lazenby et al. (2015) found both lower receptive and expressive language scores at 12 months in later diagnosed infants, and that these language patterns differentiated infants who developed ASD and those who did not.

As demonstrated in Chapter 2, a fundamental difficulty in understanding symptom onset in ASD is the variability exhibited between individuals across numerous cognitive and behavioural domains. Whilst it has been argued that early atypicalities in language domains could be a predictor for ASD later in development, the absence of delay in some infants has

also been reported. For example, after assessing language comprehension in high-risk infants, Hudry et al. (2013) reported significant differences between high-risk infants who went on to have ASD and infants who were diagnosed as typically developing. However, differences were not found between individuals with ASD and individuals who exhibited atypical behaviours but were not diagnosed with ASD (individuals classified as atypically developing either scored above the threshold for ASD in the ADOS, or scored below the average in Mullen expressive language, receptive language or early learning composite domains, but did not meet the diagnostic criteria for ASD). This result suggests that difficulties in language comprehension were not specific to infants who were later diagnosed with ASD, but were also identified in children with milder developmental delays or impairments.

Using data from a high-risk infant cohort, Landa et al. (2007) identified differences in onset patterns and levels of expressive language. Infants exhibiting ASD symptoms at 14 months produced fewer consonants in syllables than low-risk groups and infants who were diagnosed with ASD at 36 months but did not demonstrate atypical behaviours at 14 months. Identifying specific onset patterns could be essential in understanding some of the variability in the expression of ASD.

3.2.3 Executive function

Executive functioning refers to a set of processes that utilise self-control and regulation in order to organise and act on information (Hill, 2004). Currently, results from prospective, high-risk studies are limited in this domain. However, measures of visual attention and, in particular, disengagement of attention have offered insight into impairments of some aspects of executive function (Jones et al., 2014). There has been cautious support for an attention

driven theory of ASD, with the suggestion that difficulties in disengagement could lead to a focus on insignificant aspects of visual stimuli (e.g. focusing on non-social stimuli over a human face). This in turn could cause infants to disregard social cues from socially relevant stimuli (Bryson et al., 2004; Dawson et al., 2005).

Attentional disengagement has been investigated in high-risk infants using the gap overlap task. The task measures the cost of disengaging from a central stimulus in order to fixate on a peripheral visual target. A central animation stimulus is presented to orient an infant's gaze. This is followed by a peripheral target, which is presented either alongside the central stimulus (an overlap trial) or individually (the baseline condition) in order to measure the time it takes an infant to disengage from the central to peripheral stimuli. Elison et al., (2013) reported slower latencies to shift attention in overlap trials at 7 months in infants with higher ADOS scores (and therefore demonstrating a higher number of ASD-related behaviours) at 24 months. Overlap and trials were presented either with or without a temporal gap between the offset of the central stimulus and the onset of the peripheral target. Typically developing infants would, on average, orient faster after being cued by the temporal gap preceding the onset of the peripheral stimulus. This is known as the facilitation effect. Infants who went on to have higher ADOS scores (more severe autistic-like behaviours) at a 24-month assessment exhibited slower facilitation effects at 7 months. Furthermore, Elsabbagh et al. (2009) compared performances of 9-and 10-month old high-risk and low-risk infants, demonstrating longer disengagement latencies and lower facilitation effects in comparison with low-risk infants.

Two studies have looked at disengagement longitudinally, following infants across the first 36 months of life. Zwaigenbaum et al. (2005) reported difficulties with disengagement

between 9 and 12 months in infants who were later diagnosed with ASD at 24 months in comparison with both high-risk and low-risk infants. Similarly, Elsabbagh et al. (2013a) reported that whilst disengagement at 7 months was not associated with later diagnostic outcomes, individuals who were later diagnosed with ASD demonstrated longer disengagement latencies at 14 months. Furthermore, latencies increased from 7 to 14 months in individuals later diagnosed with ASD, but decreased (as one would expect) in typically developing children. Chawarska et al. (2012) reported that typically developing infants took longer to disengage from socially relevant stimuli such as human faces, whereas infants diagnosed with ASD did not show such a bias. In a more recent study, Chawarska, Macari, Powell, DiNicola and Shic (2015) identified gender differences in social attention in high-risk infants. The authors analysed responses to social engagement tasks in infants between 6 and 12 months and reported enhanced attention to faces and other social targets in high-risk females in comparison with both high-risk males and low-risk males and females. The authors suggest the possibility of an alternate expression of autism risk in females. Increased attention towards social stimuli was associated with less severe social impairments, and it was hypothesised that some high-risk females may use enhanced social attention as a protective factor against ASD. Whilst the results were preliminary, it is an exciting first step towards elucidating whether sex differences can be identified in high-risk infants.

Some developmental theories (e.g. Osterling & Dawson, 1994; Landry & Bryson, 2004) have proposed that atypical disengagement could negatively impact on multiple areas of development (e.g. joint attention, social orienting), which could lead to downstream atypicalities across a number of domains, and turn could contribute to the development of ASD symptoms. Such theories have suggested that atypical behaviours in visual attention would be some of the primary emerging symptoms. However, the majority of prospective

studies have not found evidence to support this claim. For example, Elsabbagh et al. (2013b) found that disengagement and visual attention problems were not evident earlier than other atypicalities (see also Gliga, Jones, Bedford Charman & Johnson, 2014).

3.2.4 Motor predictors

Whilst impairments in motor domains have until recently been considered by many as atypicalities which occur as a result of the emergence of initial atypical social processes, a number of studies have identified motor problems as potentially significant features in early development of ASD (Leonard et al., 2013). Several studies have identified motor atypicalities in high-risk infants who are later diagnosed with an ASD diagnosis, and have suggested potential markers of risk (Landa & Garrett-Mayer, 2006; Flanagan et al., 2010; Iverson & Wozniak, 2007; Leonard et al., 2013; Estes et al., 2015).

Landa and Garrett-Mayer (2006) found that high-risk siblings with ASD demonstrated a lower proficiency in gross motor, fine motor, receptive language and expressive language Mullen assessments in comparison with typically developing high-risk infants. Moreover, ASD-siblings displayed lower scores on all subscales in comparison with infants with a language delay at 24 months. Iverson and Wozniak (2007) compared motor and communicative development in high-risk infant siblings and low-risk infant controls between 5 and 14 months. Parents videotaped infants at home once a month, and infants were assessed at an 18-month follow-up. Overall, high-risk infants exhibited delays in the onset of developmental milestones and postural instability. The age range at which high-risk infants began walking was, on average, later than low-risk infants (between 10 and 18 months and 9 and 14 months, respectively). The authors argue that whilst findings were not specific to

infants who later developed ASD, they were specific to high-risk infants, which could be indicative of general risk status and essential in developmental screenings for ASD.

Leonard et al. (2014) identified motor differences at an earlier age in high-risk infants. Using parental interviews and a motor task battery, significantly lower motor scores were identified at 7 months in high-risk infants. Furthermore, high-risk infants later diagnosed with ASD exhibited poorer gross motor skills at 36 months in comparison with the at-risk infants who did not display developmental atypicalities. However, differences between high-risk ASD, high-risk atypical and high-risk typically developing groups were not evident at 7 months. This lack of differentiation between groups suggests that motor atypicalities are not necessarily a core feature of ASD, but perhaps a general risk factor for atypical development. Furthermore, Estes et al. (2015) recruited 210 high-risk infants and compared behaviours at 6, 12 and 24 months using Vineland, Mullen and AOSI scores. Differences were demonstrated at 6 months in gross motor and visual reception domains in high-risk infants who went on to develop and ASD diagnosis at 24 months.

Thus far, prospective studies of early infant development have identified atypicalities in executive functions, social, communicative and motor domains in high-risk infants who are later diagnosed with ASD. A number of theories suggest that the earliest atypicalities predictive of an ASD outcome later in development would be social. However, the findings reported in prospective studies have not supported this hypothesis. The vast majority of prospective studies have shown the earliest predictive differences at 12 months, not 6 months as was originally proposed. Development in infants later diagnosed with ASD appears to be indistinguishable from typically developing infants across many domains in the first 6 months of infancy. Where earlier impairments have been identified, such as in motor

domains, atypicalities do not appear to be ASD specific. Instead it is likely that atypicalities are indicative of general delay and atypical development in some individuals, and in others it may be associated with social and cognitive difficulties specific to ASD (Leonard et al., 2014). This could also support the idea of the broader autism phenotype, where high-risk individuals demonstrate elevated levels of subclinical atypicalities but do not have ASD (Ozonoff et al., 2014). The co-occurrence of difficulties across a range of domains suggests that analysing multiple scores from early developmental time points could provide an opportunity to better predict clinical outcomes.

3.3 Predictions from the Over-Pruning computational account of ASD

The Over-Pruning hypothesis of ASD (Thomas, Davis, Karmiloff-Smith, Knowland & Charman, 2015) (see chapter Chapters 4 and 5 for current research) is based on a computational model, proposing that ASD is the result of over-aggressive synaptic pruning in the first years of infancy. In the model, it is hypothesised that early, late and regressive subtypes of ASD can be produced by one pathological mechanism, an over-pruning parameter (also see Chapter 4). It also predicted that an interaction between this mechanism and population-wide variation can produce multiple profiles resembling ASD subgroups, and that unaffected high-risk siblings could either be inheriting a milder form of the over pruning mechanism, or could inherit other risk factors without the pathological pruning mechanism itself.

The model produced a number of predictions in terms of the types of atypical behaviours demonstrated in ASD, and which symptoms would be visible at the earliest time points. It is known that the onset of typical pruning in the human brain commences at specific and distinctive time points across different brain regions (Huttenlocher & Dabholkar, 1997;

Huttenlocher, 2002). In general, pruning is observed first in sensory and motor areas, followed by higher order association areas that are involved in speech production and comprehension (e.g. Wernicke's area and Broca's area), and last in the prefrontal cortex (Thomas et al., 2015). Therefore, if pruning was elevated to atypical levels, the primary emerging symptoms should be in sensory or motor rather than social communication domains. Thus, infants may exhibit early typical development in social and communicative domains while already exhibiting subtle signs of atypicality or impairment in sensory and motor domains. A second prediction from the computational account was that the first few months of infancy would be indistinguishable from typically developing infants, as pruning would not have yet commenced. A final prediction was that atypicalities would be widespread across multiple domains after the first year of infancy, and that variability resulted from population-wide variation. Individuals may demonstrate different patterns of strengths in cognitive and behavioural domains due to early experiences. Therefore, protective factors would vary amongst individuals.

Prospective research does, so far, tentatively support hypotheses from the Over-Pruning account. The first year of infancy is seen as relatively typical and early motor impairments have been displayed in at-risk infants. Furthermore, as discussed in chapter 2, the heterogeneity demonstrated in ASD has caused difficulties in fully identifying early markers predictive of later ASD outcomes. In the current chapter, I clarify theoretical questions by producing models of cumulative risk. I use behavioural data from 7 and 14 months to predict developmental outcomes at 24 and 36 months. I then test the validity of the predictive models by incorporating coefficients from significant regression models in Phase 1 to a model that applied data from an independent cohort of infants from the BASIS, known as Phase 2. In this model, outcome is not known prior to the model's production of outcome

predictions. This is the first statistical model to use data from two independent prospective datasets to validate significant findings.

3.4 Methods

3.4.1Sample

Data were made available from the British Autism Study of Infant Siblings (BASIS: http://www.basisnetwork.org/). Two cohorts of infants were recruited in the current study. The first (here-after, Phase 1) recruited 104 infants (consisting of 54 at-risk siblings and 50 low-risk siblings). Infants were assessed at around 7 months, 14 months, 24 months and 36 months. The high-risk group was defined as having an older sibling with a clinical ASD diagnosis that was confirmed by two expert clinicians. Infants in the low-risk group were recruited from a separate volunteer database from the Centre for Brain and Cognitive Development. Inclusion criteria for the low-risk group included having at least one older sibling scoring below the threshold for ASD as measured by the Social Communication Questionnaire (SCQ; Rutter, Bailey and Lord, 2003), and parental confirmation that ASD was not present in any first-degree relatives (see chapter 2 for further sample and recruitment details).

A second cohort (hereafter, Phase 2) recruited 131 infants and comprised of 105 high-risk infants and 26 low-risk infants. The criteria for inclusion in both groups were identical to Phase 1. All infants in Phase 2 were assessed at around 5 months, 8 months, 14 months, 24 months and 36 months. In the current chapter, data from time points 3 (14 months) and 4 (24 months) were analysed.

3.4.2 Measures

In order to measure developmental levels, infants were assessed on one standardised test battery, the *Mullen Scales of Early Learning* (MSEL; Mullen, 1995). A parental report, the Vineland Adaptive Behaviour Scales (*VABS)* was also used as a second measure of infant development (see Chapter 2 for a full description of assessments and for Phase 1 descriptive statistics). Atypical behaviours that were characteristic of ASD were measured using the *Autism Observation Scale for Infants* (AOSI; Bryson, Zwaigenbaum, McDermott, Rombough, and Brian, 2008).

Mullen Scales of Early Learning

The Mullen Scale of Early Learning is a standardised developmental measure used to assess early cognitive and motor development (see chapter 2 for full description). Data from 14-month visits were utilised in the current chapter. Data from the five subscales was assessed: fine motor, gross motor, visual reception, receptive language and expressive language. Table 3.1 provides descriptive statistics for 7-and 14-month T-scores from Phase 1 and 14-month T-scores from Phase 2.

Table 3.1. Mean and standard deviation statistics for the MSEL T scores at 7 and 14 months from Phase 1 and 14 months from Phase 2 data collection.

Phase 1	**Low-Risk Controls**		**High-Risk Siblings (Combined)**		**Below ASD Cut-off (24-month ADOS scores)**		**Above ASD Cut-off (24-month ADOS scores)**	
7 Months								
Mullen Gross Motor	50.40	9.09	45.4	9.99	45.33	8.84	46.06	12.58
Mullen Visual Reception	54.32	8.66	50.53	8.96	50.67	7.86	50.00	11.49
Mullen Fine Motor	58.00	9.41	52.45	10.46	53.67	10.26	49.81	11.08
Mullen Expressive Language	50.20	8.53	42.98	7.07	42.36	5.79	44.69	9.45
Mullen Receptive	46.04	9.03	41.42	11.79	42.89	10.01	38.81	15.08

Language

	N=50		***N=54***		***N=36***		**N=18**	
14 months								
Mullen Gross Motor	51.06	16.07	46.26	16.55	47.14	15.75	45.53	18.44
Mullen Visual Reception	55.70	9.39	51.19	10.55	53.31	10.23	47.71	10.11
Mullen Fine Motor	61.17	9.10	54.94	12.44	57.26	11.72	51.18	12.91
Mullen Expressive Language	48.08	9.36	44.42	11.46	46.66	11.03	39.88	11.69
Mullen Receptive Language	46.10	12.36	43.36	12.95	46.63	12.84	37.76	10.69
	N=50		***N=54***		***N=36***		**N=18**	
Phase 2								
14 months								
Mullen Gross Motor	50.22	14.00	49.88	13.99	51.22	13.57	45.41	14.74
Mullen Visual Reception	52.26	7.94	46.24	9.16	47.61	8.99	43.16	9.26
Mullen Fine Motor	54.33	6.97	53.64	9.01	55.12	7.16	48.62	12.46
Mullen Expressive Language	53.23	11.01	48.29	10.55	49.69	9.66	43.58	12.18
Mullen Receptive Language	45.88	10.55	40.18	11.17	41.98	11.19	34.08	8.85
	N=26		***N=105***		***N=80***		**N=25**	

Phase 1

7 months

On average, low-risk infants demonstrated significantly higher abilities than high-risk infants in gross motor $t(103) = 2.65$, $p = .009$, and expressive language $t(103) = 4.68$, $p = .001$.

A comparison between high-risk infants below the ASD cut-off and infants above the ASD cut-off revealed no significant differences between scores at 7 months.

14 months

Low-risk infants demonstrated significantly higher abilities than high-risk infants in visual reception $t(103) = 2.4$, $p = .027$ and fine motor $t(103) = 2.82$, $p = .006$.

A comparison between high-risk infants below the ASD cut-off and high-risk infants above the cut-off for ASD revealed significant differences at 14 months. High-risk infants below the ASD cut-off scored higher in expressive language $t(53) = 1.97$, $p = .016$, and receptive language $t(53) = 2.48$, $p = .055$

Phase 2

On average, low-risk infants demonstrated significantly higher abilities than high-risk infants in visual reception $t(130) = 23.35$, $p = .002$, receptive language $t(130) = 2.24$, $p = .026$ and expressive language $t(130) = 2.06$, $p = .036$ at 14 months.

A comparison between high-risk infants below the ASD cut-off and high-risk infants above the cut-off for ASD revealed differences at 14 months. High-risk infants below the cut-off for ASD scored higher in fine motor $t(104) = 3.24$ $p = .021$, receptive language $t(104) = 3.17$, $p = .002$ and expressive language $t(104) = 2.55$, $p = .031$. Table 3.1 displays the descriptive statistics for all groups.

3.4.3 Vineland Adaptive Behaviour Scales (VABS)

Data from the Vineland Adaptive Behaviour Scales were also used as a second measure of infant development. Table 3.2 provides descriptive statistics for 7-month and 14-month scores in Phase 1, and 14-month scores in Phase 2.

3.2. Mean and standard deviation statistics for the Vineland Standard scores at 7 and 14 months from Phase 1 and 14 months from Phase 2 data collection.

Phase 1	**Low-Risk Controls**	**High-Risk Siblings (Combined)**	**Below ASD Cut- off (24-month ADOS scores)**	**Above ASD Cut- off (24-month ADOS scores)**

7 months								
Vineland Communication	103.33	14.33	93.11	18.18	96.2	13.23	89.35	12.28
Vineland Motor	98.80	14.34	85.25	16.26	84.94	15.79	85.53	18.08
Vineland Daily living	100.57	16.59	97.92	17.65	99.63	18.17	93.29	16.09
Vineland Socialisation	104.24	12.61	97.74	14.55	97.94	14.94	97.71	14.52
	N=50		***N=54***		***N=36***		***N=18***	
14 months								
Vineland Communication	101.11	8.39	92.06	16.08	95.12	16.94	86.59	15.96
Vineland Motor	104.76	12.02	94.16	14.01	93.73	12.97	94.94	16.59
Vineland Daily living	98.11	9.99	90.01	13.76	92.61	12.75	85.82	15.09
Vineland Socialisation	100.35	10.24	95.37	14.22	98.12	13.75	90.24	14.48
	N=50		***N=54***		***N=36***		**N=18**	
Phase 2								
14 months								
Vineland Communication	101.50	12.20	95.73	12.79	97.67	10.63	93.95	11.69
Vineland Motor	102.80	9.23	100.48	12.29	101.07	11.46	98.19	15.20
Vineland Daily living	100.65	12.91	95.17	12.79	96.74	12.92	89.14	14.87
Vineland Socialisation	100.82	12.32	96.98	10.91	9 7.76	10.64	93.95	11.68
	N=26		***N=105***		***N=80***		**N=25**	

Phase 1

7 months

Low-risk infants displayed higher ability scores than high-risk infants in the domains of communication $t(103) = 3.13$, $p = .002$ and socialisation $t(103) = 2.40$, $p = .018$. When comparing high-risk groups, scores were not significantly different between groups in any domain.

14 months

Low-risk infants demonstrated significantly higher ability scores in communication domains

$t(103) = 3.32$, $p = .001$ and motor domains $t(103) = 3.98$, $p = .001$. When comparing high-risk groups, scores were not significantly different between groups in any domain.

Phase 2

Low-risk infants only displayed higher ability scores than high-risk infants in one domain, communication $t(130) = 2.06$, $p = .040$, at 14 months. However, when comparing high-risk groups, scores in communication $t(104) = 3.16$ $p = .012$ and daily living domains $t(104) = 3.29$ $p = .022$ were significantly larger in the high-risk below the ASD cut-off group when compared with individuals in the high-risk above ASD cut-off group.

3.4.4 Autism Observation Scale for Infants

The *Autism Observation Scale for Infants* (AOSI; Bryson, Zwaigenbaum, McDermott, Rombough, & Brian, 2008) is a semi-structured observational assessment developed to measure and detect early signs of ASD-related behaviours in high-risk infants between 6 and 18 months. Observations from semi-structured play sessions are used to code 19 items, and each item is coded on a scale between 0 and 2 (See Brian et al., 2008 for a full description of all items). In the current study, a total score (total of all 19 points, with a maximum score of 44) was generated, where a higher score was indicative of higher levels of atypicality. The majority of assessments were double-coded by an examiner and observer with high inter-rater reliability in Phase 1 ($n = 85$, intraclass correlation coefficient = .95) and Phase 2 ($n=115$, intraclass correlation coefficient = .96). See table 3.3 for 7- and 14-month descriptive statistics from phase 1 and 14-month descriptive statistics from Phase 2.

Table 3.3. Mean and standard deviation statistics for AOSI Total scores at 7 and 14 months from Phase 1 and 14 months from Phase 2 data collection.

Phase 1	**Low-Risk Controls**		**High-Risk Siblings (Combined)**		**Below ASD Cut-off (24-month ADOS scores)**		**Above ASD Cut-off (24-month ADOS scores)**	
7 months								
AOSI Total Score	7.12	4.07	9.33	5.63	8.83	5.14	10.82	6.42
	N=50		***N=54***		***N=36***		***N=18***	
14 months								
AOSI Total Score	3.17	3.25	4.64	4.47	3.97	4.38	6.18	4.50
	N=50		***N=54***		***N=36***		***N=18***	
Phase 2								
14 months								
AOSI Total Score	4.07	3.58	6.31	4.71	5.40	4.08	9.41	5.43
	N=26		***N=105***		***N=80***		***N=25***	

Phase 1

7 months

As a group, high-risk infant AOSI total scores were higher than low-risk infants $t(103) = 2.31$, $p = .026$. Scores were not significantly different between high-risk subgroups.

14 months

High-risk infant AOSI total scores were significantly higher than low-risk infants $t(103) = 2.98$, $p = .023$. When comparing high-risk subgroups, the high-risk infants above the cut-off for ASD demonstrated a higher mean total score in comparison with high-risk infants below the ASD cut-off $t(53) = 2.31$, $p = .025$.

Phase 2

Independent t-tests showed that as a group, high-risk infant AOSI total scores were higher than low-risk infants $t(130) = 2.23$ $p = .025$. High-risk infants above the cut-off for ASD demonstrated a higher mean total score in comparison with high-risk typically developing infants $t(104) = 3.91$ $p = <.005$).

3.4.5 Parental Demographics

Parental interviews generated demographic information for each family. The central factors were overall household income, current job title and highest academic qualification for both parents, and the number of whole and half siblings in the family. Overall household income was assessed using an ordinal scale from 1-5, where 1 was "earning less than £20,000 per year" and increasing in increments of £20,000, where 5 was "earning over £80,000 per year". Factors relating to job title and qualifications were transformed to an ordinal scale to allow for comparison between measures. Parental education levels were rated on a scale of 1-5, 1 defined as "finishing school with qualifications up to GCSE level", and 5 being "above Masters level or equivalent" training. Job type was rated on a 1-5 scale. In this case, job type was defined in terms of salary, where 1 referred to stay-at-home parents or retired individuals, 2 referred to part-time workers, 3 4 and 5 referred to lower, average and high paying job-type respectively. All category values were independently blind-rated and Cohen's k was used to assess interrater reliability. Agreement between the two judgments was excellent k = .931 (95% CI, .864 to .972, $p < .005$). When codes differed between raters, they agreed on a consensus code. See Appendix 1 for full demographics table.

3.4.6 Diagnostic outcome

Outcome at 24 months

At 24 months, all high-risk infants participated in a preliminary diagnostic evaluation. In

Phase 1, this evaluation was based solely on Module 1 ADOS scores. All infants were assessed using the Autism Diagnostic Observation Schedule Generic (ADOS-G; Lord et al., 2000) in Phase 1 and in Phase 2. This is a semi-structured measure used to assess social and communicative skills related to diagnosing ASD. Such skills include reciprocal social interaction, play, eye contact, gestures and stereotyped behaviours and restricted interests. Behaviours are coded and scored using a scale from 0 to 3, with a higher score indicating a greater level of autistic-like atypicality. In the current chapter, the ADOS-G ASD algorithm (a single composite score based on the sum of all scores across tasks) was used. Scores were compared with the algorithm cut-off score for ASD and autism. Low scores (below 13) were indicative of typical development. A score between 13 and 16 categorised an infant as above the cut-off point for ASD, while a score of 16 or above categorised an infant as being above the cut-off point for autism and was indicative of a higher number of ASD related behaviours and a higher severity of symptoms. Here, all statistical models that use diagnostic outcomes at 24 months are derived from the ADOS total score cut-off point. High-risk infants were categorised as either below the cutoff for ASD, or above the cutoff for ASD.

Phase 1

In Phase 1, out 54 high-risk infants, a total of 36 toddlers (21 females, 15 males) were categorised as high-risk typically developing and 18 (11 females, 6 males) were categorised as above the threshold for ASD (representing 33.3% of the high-risk group).

Phase 2

In Phase 2, out of 105 high-risk toddlers, 80 (41 females, 39 males) were categorised as high-risk typically developing, and 24 (7 females, 17 males) were categorised as above the ADOS cut-off for ASD (representing 28.9% of the high-risk group).

Outcome at 36 months

Phase 1

Behavioural and environmental measures were also used as predictors of Phase 1, 36-month outcomes. Here, diagnostic outcomes were used. Outcome at 36 months combined data from all four infant visits and assessments by a team of clinicians in order to establish a diagnostic classification. The data used for diagnostic evaluation included results from both 24 and 36-month visits, including scores from the *ASD Diagnostic Interview-Revised* (ADI-R; Lord et al., 1994). This was administered in the form of semi-structured interview and was compared with scores from the Autism Diagnostic Observation Schedule-Generic (*ADOS-G:* Lord et al., 2000).

As in Chapter 2, toddlers in high-risk and high-risk atypical categories were collapsed to form two outcome groups: high-risk non-ASD and high-risk ASD. A total of 53 infants were assessed at 36 months. Under these classifications, 36 toddlers (26 females, 10 males) were classified as high-risk no ASD and 17 (6 females, 11 males) as high-were categorised as high-risk ASD.

Phase 2

In the Phase 2 cohort, 36-month diagnostic outcomes had not been collected at the time of analysing the current data. Therefore, models using 36-month outcomes only utilised data from Phase 1. Validation analyses using Phase 2 data were only run based on 24-month ADOS classification measures.

3.4.7 Data analysis

Behavioural and environmental measures from 14 months were used as predictors of

infant outcome at 24 and 36 months. The first part of the results utilises data from Phase 1. In order to select factors that may be predictive of later outcomes, mixed ANOVAs were utilised to compare 7- and 14-month Mullen and Vineland scores with 36-month outcome groups. Measures with significant differences at 7 and 14 months were retained for further analyses. To avoid multicollinearity between measures, correlations were run on the remaining behavioural and cognitive assessments, and correlated predictors were removed from analyses. In order to include environmental data as a predictive factor, a principal component analysis (PCA) was run on parental data relating to socio economic status (SES). These factors were total household income, mother's highest qualification and job title, and father's highest qualification and job title. The PCA produced two environmental factors. The predictive value of each non-correlated parameter at 7 months and 14 months was then assessed against outcome on ADOS scores at 24-months to identify significant behavioural predictors, and at 36 months to compare the stability of predictive factors.

The second part of the results utilised data from Phase 2. In order to test the effectiveness of significant regression models from Phase 1, the regression models were applied to a second dataset, consisting of identical variables taken from Phase 2 BASIS data cohort. Based on the coefficients from the significant Phase 1 regression model, outcome predictions were produced for the Phase 2 data. In order to examine the consistency of prediction rate between the two phases, results between the two cohorts are presented.

3.5 Results

3.5.1 Phase 1

Exploratory analyses for predictors of outcome

In order to investigate the relationship between group outcome and early cognitive ability, Mullen and Vineland raw scores were entered into mixed ANOVA models. One cognitive subdomain score was compared in each model across the first two time points at which at which infants were assessed (7 and 14 months). This was tested against outcome group, which consisted of four levels. These were low-risk, high-risk, high-risk below the ASD cut-off and high-risk below the ASD cut-off. Whilst all statistical predictive models only include high-risk infants, it was beneficial to compare scores on behavioural metrics across all groups to assess whether development was in the typical range at 7 and 14 months. Interaction effects between age, cognitive ability and outcome grouping are discussed below.

Mullen Comparisons

Each ANOVA included ability scores from one of the five Mullen sub-domains (gross motor, visual reception, fine motor, receptive language and expressive language) as the dependent variables. Outcome group was used as the between-group variable and contained four levels. Time point was the within-subject variable, which contained two levels, 7 months (time point one) and 14 months (time point two).

Prior to analyses, box plots were used to identify outliers in the data for all sub domain and time points. Outliers were identified by inspection of boxplots greater than 1.5 box lengths from the edge of the box. A total of four outliers were identified, and were between 1.5 and 2.5 box lengths from the edge of the box. Rather than removing the outliers, each case was modified by replacing the extreme score with the value of next largest non-outlier (Pearson, 2002). As scores for all subdomains were normally distributed as measured by the Shapiro-Wilk's test ($p < .005$) a transformation of the outliers was considered unsuitable.

Gross motor

No significant interaction was identified between outcome group and time point on gross motor scores. *F(3,88)* = 2.953, *p* = .120, partial η^2 = .168.

Visual reception

No significant interaction was identified between outcome group and time point using visual reception scores *F(3,88)* = 1.951, *p* = .133, partial η^2 = .172.

Fine motor

Homogeneity of variances was demonstrated using Levene's test (*p* > .05). Homogeneity of covariance was also shown using Box's test of equality of covariance matrices (*p* = . 210). There was a statistically significant interaction between outcome group and time point on fine motor scores *F(*3,88) = 6.090, *p* = .003 partial η^2 = .110. Differences between outcome groups at 7 months were non-significant *F*(3,91) = 2.992, *p* = .052, partial η^2 = .090. A significant difference in fine motor scores was identified between outcome groups at 14 months *F*(3,88*)* = 1.951, *p* = .006, partial η^2= .067. Fine motor scores were significantly lower in the high-risk below ASD threshold group (M = -2.27, SE = .632, *p* = .010) and the high-risk above ASD threshold group (M = -.172, SE = .337, *p* = .001) when compared with the low-risk group.

Receptive language

Homogeneity of variances was demonstrated using Levene's test (*p* > .05). There was no homogeneity of covariance, as shown using Box's test of equality of covariance matrices (*p* = .023), however the ANOVA was run as this analysis is reasonably resilient to homogeneity of covariance violations. There was a statistically significant interaction between outcome group and time point on receptive language scores *F(*=3,91) = 2.172, *p* = .024, partial η^2 =

.071. There was a significant difference in receptive language scores between outcome groups at 14 months $F(3,91) = 4.614$, $p = .002$, partial $\eta^2 = .132$. Receptive language scores were significantly lower in the high-risk above ASD threshold group (M = -1.67, SE = .519, $p = .009$) when compared with the low-risk group. Scores were not significantly different in the high-risk above ASD threshold group (M = -.1.38, SE = .571, $p = .081$) or the high-risk below ASD threshold group (M = -.74, SE = .434, $p = .321$).

Expressive language

Homogeneity of variances was demonstrated using Levene's test ($p > .05$). There was also homogeneity of covariance, as shown using Box's test of equality of covariance matrices ($p = .098$). A statistically significant interaction was identified between outcome group and time point on expressive language scores $F(6,95) = 6.177$, $p = .002$, partial $\eta^2 = .110$. There was a significant difference in expressive language scores between outcome groups at 14 months $F(3,95) = 3.686$, $p = .015$, partial $\eta^2 = .114$. Expressive language scores were significantly lower in the high-risk above ASD threshold group (M = -5.21, SE = 1.636, $p = .011$) when compared with the low-risk group. Scores were not significantly different in the high-risk below ASD threshold group (M = -.1.06, SE = .1.634, $p = .883$).

Vineland scores

Interactions between outcome groups at 7- and 14-months were non-significant across all Vineland domains. These were communication scores $F(3,96) = .479$. $p = .698$, partial $\eta^2 = .016$, daily living scores $F(3,96) = .566$, $p = .639$, partial $\eta^2 = .019$, socialisation scores $F(3,96) = .818$, $p = .487$, partial $\eta^2 = .027$ and motor scores $F(3,96) = .916$, $p = .164$, partial $\eta^2 = .134$.

Environmental measures

In order to include environmental effects as a predictor of outcome, seven environmental measures from the BASIS were included in a principal component analysis (PCA). A PCA was selected as the most appropriate variable reduction method. Creating an uncorrelated and linear combination of weighted observed variables could explain the maximal amount of variance explained in comparison with summing the scores of all measured variables. All measures were collected from parental interviews. The factors were mother's highest qualification, father's highest qualification, overall household income, mother's job, father's job, number of whole siblings and number of half siblings (see the Methods section for further detail).

A PCA was run on the seven parental report measures. The correlation matrix showed that two measures had correlation coefficients of less than .03 (number of whole siblings and number of half siblings). Therefore, both were removed from further analyses. The PCA was run with the remaining five measures. Overall, the Kaiser-Meyer-Olkin (KMO) measure was .531 with individual KMO measures all above the accepted value of .05. Bartlett's test of sphericity was also significant ($p <$. 005) which demonstrated that no variables were correlated, and could therefore be factorised. In order to decide which components should be retained, the eigenvalue-one criterion was used. Two components with eigenvalues greater than one were identified. Component one explained 38.5% and component two explained 25% of the total variance. A Varimax orthogonal rotation was employed to aid interpretability. The rotation solution exhibited 'simple structure' (Thurstone, 1947).

Table 3.4 Rotated structure Matrix for PCA with Varimiax rotation with two components.

Items	Rotated Component Coefficients	
	Component 1	Component 2
Fathers job	0.854	
Overall household income (1-5)	0.741	
Father highest Qualification	0.713	
Mother highest qualification		0.859
Mothers job		0.707

The two-component solution explained a 63.5% of the total variance; therefore the other three components were not included. The interpretation of the data was consistent with attributes from the parental interviews. Component 1 comprised of overall house income, father's highest qualification and job type. Component 2 was comprised of mother's job and highest qualification. Component loadings of the rotated solution are presented in Table 3.4.

Phase 1 regression analyses

Using 7-month data to predict 24-month outcome

In order to explore the predictive power of social, communicative, sensory and motor measures, binary logistic and multiple linear regression models were utilised. Initially, scores from the earliest time point (7 months) were used to predict group membership at 24 months. However, all regression models were non-significant. Based on the exploratory ANOVAs, this was as expected.

Using 14-month data to predict outcome

Models used behavioural and cognitive measures from 14 months as predictors for outcome at 24 and 36 months. Binary logistic regressions were used here. At 24 months, logistic regression models predicted classification into one of two group outcomes: high-risk below the threshold for ASD, or above the threshold for ASD. Individuals scoring above the

threshold for ASD or above the threshold for autism on the ADOS were collapsed into one group in order to retain a larger sample size. At 36 months, logistic regression models classified individuals as either high-risk non-ASD or high-risk ASD. The high-risk group below the threshold for ASD included both high-risk typically developing and high-risk 'other' individuals.

At 24 months, total ADOS scores were used as the outcome measure in linear regression models. A lower ADOS score (defined by the ADOS-G algorithm score and using the ASD threshold of 13 or lower) indicated typical development, and a higher score (above 13) indicated that behaviours indicative of ASD or autism were present. At 36 months, diagnostic outcomes were defined using a combination of clinical assessments, and therefore did not lie on a continuum. Rather, scores were categorical in origin, and therefore logistic regression analyses were appropriate at this time-point.

24m regression models

Binary logistic regression

A binary logistic regression was run to ascertain the effects of Mullen Fine Motor, Expressive Language, AOSI total score and Environmental scores on the likelihood that individuals would be classified as above or below the threshold for ASD. Linearity was confirmed using the Box-Tidwell procedure. All independent variables were linearly related to the logit of the dependent variable (p = <.005). The regression was statistically significant X^2 (5) = 18.469, p <.005. The model explained 49.2% Nagelkerke R^2 of the variance in ADOS outcome, and correctly classified 82.9% of the cases overall. Sensitivity was 69% and specificity was 92%. Positive prediction value (those above the ASD cut-off) showed an 84.6% correct prediction rate. Negative Prediction Value (those below ASD cut-off) showed an 82.1% correct

prediction rate. The statistically significant predictors were Mullen expressive language, p = .010 and environmental factor 1, p = .04. Table 3.5 shows the significant predictors of ASD. A lower expressive language and lower environmental score (consisting of lower house income and lower parental education levels) were associated with an increased possibility that an individual would be categorised above the ADOS cut-off for ASD or autism.

Table 3.5. Significant predictors of outcome in the binary logistic model.

	B	SE	Wald	*df*	*P*	Odds Ratio	Lower	Upper
Mullen Expressive Language	-0.16	0.062	6.577	1	0.01	0.852	0.754	0.963
Environmental Factor 1	-0.935	0.465	4.051	1	0.044	0.393	0.158	0.976

Multiple Linear regression

A multiple linear regression was run to predict 24-month ADOS scores from Mullen fine motor, gross motor, expressive language, visual reception, AOSI total score and the two environmental factors (environmental factor 1 and environmental factor 2). The model significantly predicted ADOS score, F (5,35) = 4.08, p < .005, R^2 adjusted = *.29*. As with the binary logistic regression, two variables significantly added to the model, expressive language, p <.005 and Environmental Factor 1, p = .035. All other predictors were above .248. Table 3.6 shows the significant predictors.

Table 3.6 significant predictors of outcome from the Phase 1 multiple linear regression model.

Variable	B	*Seβ*	*β*
Mullen Expressive Language	-0.211	0.074	-0.394	-2.831
Environmental Factor 1	-1.937	0.885	-0.296	-2.19

Analyses using Phase 1 data were exploratory as infant outcomes were known prior to the onset of analyses, and used to build the predictor models. Using Phase 2 data allowed a priori testing without prior knowledge of individual outcomes, and therefore verification of the Phase 1 regression models. I first looked at non-significant factors to determine their suitability for the model. Environmental factor 2 (consisting of mother's job title and income) and receptive language, when removed did not alter the outcome of the models, and so were removed from further analysis. However, AOSI total score and fine motor, whilst not significant, did improve the outcome of the model and were retained for further analyses. It is likely that these factors were involved in higher order interactions and changed the interactions of the higher order terms in the model when they were removed. It is also possible that these measures were acting as suppressor variables, whereby their inclusion in the model strengthened the effects of expressive language and environment factors.

Comparisons with 36-month outcomes

First, 24-month outcomes were tested from Phase 1 in order to directly compare predictive results with Phase 2 data, as 36-month outcome data had not been collected and verified in the Phase 2 cohort. In order to explore the stability of cognitive and behavioural domains across time and whether measures predicting outcome at 24 months were still significant at a later point in development, an additional regression model was run. Phase 1 measures were used as predictors of outcome at 36-months.

The binary logistic regression significantly predicted ADOS outcome (high-risk ASD or high-risk non-ASD) at 36 months X^2 (5) = 12.949, p = .006. The model explained 44% Nagelkerke R^2 of the variance in ADOS outcome, and correctly classified 79.4% of the cases overall. Sensitivity was 64% and specificity was 90%. Positive prediction value (those above the ASD cut-off) showed a 50% correct prediction rate. Negative Prediction Value (those diagnosed as having ASD) showed an 80% correct prediction rate.

One variable, AOSI total score significantly added to the model. A higher AOSI total score was associated with an increased likelihood of being above the typical threshold as measured by the ADOS scale. Table 3.7 shows the significant predictor values.

Table 3.7. Odds ratios for the significant predictor, AOSI total score.

						95% CI for Odds Ratio		
	B	SE	Wald	*df*	*P*	Odds Ratio	Lower	Upper
AOSI Total Score	0.322	0.143	5.085	1	0.024	1.38	1.043	1.827

Thus far, cognitive measures from the Mullen and Vineland were analysed in mixed ANOVAs to assess which measures should be included as predictors of outcome. Expressive language, receptive language and fine motor scores were retained for further analyses. A PCA was run using parental data to create two environmental predictors. Binary logistic and multiple linear regressions were run using 7- and 14-month measures as predictors of 24-

month outcomes scores. Data from 7 months was non-significant. Measures from 14 months were then used to predict outcome at 36 months to assess the stability of predictors.

3.5.2 Phase 2 results

Logistic regression models

In order to test the reliability of the significant model from Phase 1, data from Phase 2 was applied to the binary logistic regression model, predicting outcome at 24 months. Whilst both linear and logistic regression models demonstrated significant predictive power, the binary logistic regression procedure categorised individuals as having ASD or no ASD rather than predicting severity scores on a continuum. Therefore, a binary logistic regression was appropriate here. Coefficients from the Phase 1 binary logistic model were used with Phase 2 measures to compute the predicted outcome of cases of high-risk infants from Phase 2.

First, coefficients were saved from the Phase 1 binary logistic regression using Mullen fine motor, expressive language, AOSI total score and environmental factor 1 to successfully predict individuals above the threshold or below the threshold for ASD. The model was reconstructed using these coefficients and applied to the same variables, but using data from 14 months from Phase 2. The aim was to produce predicted outcomes for Phase 2 data that could be compared with actual 24-month outcomes.

In the first comparison procedure, the new model predicted individuals as below the cut-off for ASD or above the cut-off point for ASD and Autism. Table 3.8 shows the prediction results. A total of 77% of individuals were correctly classified overall. Sensitivity was 78.5%, and specificity was 81.4%. Positive prediction value (those correctly identified as

being above the cut-off for ASD or Autism) was 55.5%, and negative value (those correctly identified as being below the cut-off point for ASD) was 91.6%.

Table 3.8. Classification table showing observed versus predicted values for individuals below the cut-off for ASD and individuals above the cut-off for ASD and autism.

	ADOS Algorithm/Instrument Diagnostic Classification at 24 months		
(Observed)	Below ASD cut-off	(Predicted) Above Autism cut-off	Percentage of total correct predictions
Below ASD cut-off	58	12	81.4%
Above Autism Cut-off	2	5	71.4%
Overall percentage			81.5%

A second model was run using Phase 2 data to examine whether the model could better predict infants with 'core' autism (those above the cut-off for autism) rather than individuals who exhibited fewer symptoms and were classified as above the cut-off for ASD. In this binary logistic regression, the model predicted individuals to be either below the cut-off for ASD or above the cut-off for autism. Individuals who were categorised as above the threshold for ASD but not autism were not included in the analysis.

Table 3.9 shows the prediction results. A total of 81.5% of individuals were correctly predicted overall. Sensitivity was 71.4%, and specificity was 81.4%. Positive predictive value was 70.5%. That is, of all cases predicted as having ASD, 70.5% were correctly identified. Negative prediction value was 96%, that is, of all cases predicted as not having ASD, 96% were correctly classified.

Table 3.9. Classification table showing observed versus predicted values for individuals below the cut-off for ASD and individuals above the cut-off for ASD.

	ADOS Algorithm/Instrument Diagnostic Classification 24 months		
(Observed)	Below ASD cut-off	(Predicted) Above cut-off for ASD	Percentage of total correct predictions
Below ASD cut-off	55	15	78.5%
Above autism cut- off	5	12	70.5%
Overall percentage			77.0%

Both model iterations predicted outcome with differing degrees of accuracy. Prediction was particularly strong in the autism-only analysis, demonstrating a higher success rate when predict individuals demonstrating greater severity at outcome. The model also predicted individuals below the cut-off point for ASD more reliably, compared to individuals demonstrating atypicality.

3.6 Discussion

The current chapter aimed to determine whether scores from behavioural and environmental measures at 7 and 14 months could be applied to statistical regression models to predict high-risk infant outcomes later in infancy. To my knowledge this is the first predictive risk model in ASD that has been validated by a second independent dataset. A second aim was to clarify hypotheses arising from results of the Over-Pruning account of ASD. More specifically,

whether infants who received an ASD diagnosis demonstrated early sensory and motor atypicalities and whether atypical behaviours were widespread across behavioural domains. Statistical regression models using data from 7-month assessments did not significantly predict ADOS outcomes at 24 or 36 months. Based on predictions from the Over-Pruning account, I hypothesised that the earliest atypicalities demonstrated in infants who were later diagnosed with ASD would be in sensory or motor domains; however, the results contradict this hypothesis. The interpretation of this result is discussed below. Conversely, logistic regression models utilising data from 14 months did significantly predict ADOS outcomes at both 24 and 36 months. Specifically, expressive language, environmental factor 1, AOSI total score and fine motor scores from 14 months added significantly to the predictive ability of the models. This result provides evidence of between-group differences at 14 months that were widespread across multiple domains in infants who were later diagnosed with ASD. The results whilst preliminary support a cumulative, predictive model of ASD that incorporates multiple measures of development to predict later developmental outcomes. Furthermore, the findings were validated by an independent cohort from which the predictive markers were derived. I discuss the results and implications for future directions below.

Mixed ANOVAs were used as an exploratory method for identifying early differences in cognitive and behavioural measures that could be used to differentiate outcome groups, and the first method of identifying relevant predictors for later regression models. At 7 months, no significant differences were identified between high-risk and low-risk infants. At 14 months, significant differences were identified in fine motor scores between high-risk ASD and low-risk infants. Specifically, as a group, high-risk infants exhibited lower ability scores across fine motor measures in comparison with low-risk infants. However, there was no significant distinction between high-risk ASD and high-risk typically developing groups.

Furthermore, high-risk infants later diagnosed as above the threshold for ASD also exhibited significant differences in comparison with low-risk infants in expressive and receptive language domains at 14 months, but with no significant differences between high-risk groups. High-risk infants in general showed lower language abilities at 14 months. This lack of differentiation between high-risk infants, but significant distinction when compared with low-risk infants supports the idea of the broader autism phenotype (BAP). The BAP is a group of subclinical features that are displayed in elevated rates in family members of individuals with ASD, and is comprised of characteristics that include core diagnostic features including language delays, social problems and repetitive and rigid interests and behaviours (Ozonoff et al., 2014). The lower language and motor scores seen in high-risk infants as a group indicates that at least some of the high-risk infants below the ADOS threshold exhibit atypical development, but at a subclinical level.

Seven-month measures were first used in predictive regression models. Varied combinations of language, motor and behavioural scores were included in all iterations of the regression model, however no model showed significant predictive power. Ability scores were not significantly distinguishable between individuals in the high-risk category at 7 months, regardless of outcome categories from 24 or 36 months. The fact that at this age, cognitive ability did not correlate with later outcomes was not unexpected. Thus far, prospective research has shown few behavioural distinctions before the age of 12 months in social interaction (Rozga et al., 2011), gaze following (Young et al., 2011; Elsabbagh et al., 2013b) disengagement tasks (Bedford et al., 2012) and language domains (Zwaigenbaum et al., 2005).

Following the removal of measures that did not add predictive power to the models, both linear and logistic models significantly predicted outcome at 24 months from ability scores at 14 months using measures from the Phase 1 BASIS cohort. Linear regression models predicted the severity of ADOS scores, and the binary logistic model predicted whether individuals were above or below the threshold for ASD or autism using ADOS scores. Infants with a lower expressive language score, as measured by the Mullen expressive language scale at 14 months were more likely to be diagnosed with ASD or autism compared to high-risk infants scoring in the typical range of the ADOS assessment at 24 months. A number of prospective studies have reported comparable differences in expressive language subscale scores at 12 and 14 months in high-risk infants later diagnosed with ASD (Landa & Garrett-Mayer, 2006; Zwaigenbaum et al., 2006).

The second significant predictor in the regression models was environmental factor 1, which included overall household income, father's job category and father's highest educational level. The direction of the coefficients in the regression models showed a relationship between environmental level and diagnostic outcome. Overall, the model showed an increased likelihood for high-risk infants with a lower environmental score to be categorised as above the threshold for ASD or autism when combined with a lower expressive language score at 14 months of age.

As discussed in the introduction, the Over-Pruning computational account of ASD (see chapters 4 and 5) produced a number of hypotheses that can be directly compared here. One prediction was that the first few months of infancy would be indistinguishable between infants who are later diagnosed with ASD, and low-risk controls. The fact that no 7-month measures were predictive of later outcome supports this claim. The Over-Pruning account

also predicted early differentiation between high-risk and low-risk infants in motor domains, however this was not supported by the results in this chapter. A lack of early signs of atypicality in sensory and motor domains could provide evidence against the computational account of ASD. However, an alternative explanation for this discrepancy between the results and hypothesis could be that the Mullen and Vineland motor measures could not provide enough sensitivity to identify subtle motor atypicalities. A small number of high-risk prospective studies have provided putative evidence of early motor delays in infants who are later diagnosed with ASD (Zwaigenbaum et al., 2005; Flanagan et al., 2012). Focusing on postural control, Flanagan et al. (2012) found that infants who were later diagnosed with ASD exhibited significantly higher levels of head lag at 6 months compared to other high-risk infants or low-risk controls. Furthermore, Zwaigenbaum et al. (2005) found that parents reported lower activity levels at 6 months in infants with a later ASD diagnosis compared with other high-risk and low risk outcome groups. Whilst these results are preliminary and used small numbers of high-risk infants, they do provide evidence to support the possibility of motor atypicalities in high-risk infants later diagnosed with ASD. Leonard et al. (2014) analysed Mullen and Vineland motor scores in the Phase 1 BASIS cohort and suggest that it would be beneficial to consider more fine-grained assessments of motor skills to fully understand motor development in high-risk individuals with ASD.

A second hypothesis was that atypical behaviours would be widespread and evident across multiple cognitive domains. Significant logistic regression models showed that lower scores in language, environment and motor domains at 14 months were predictive of diagnostic outcomes at 24 months. Whilst the results from the model are preliminary, I argue that these findings support the hypothesis from the Over-Pruning account of ASD. The account predicts that after the first year in infancy, the manifestation of ASD symptoms will not be

domain specific; instead, individuals will exhibit widespread atypicalities across language, communicative, sensory and motor domains. Research examining the timing of the onset of synaptic pruning in typical development suggests that pruning will not commence in brain regions that are responsible for language, social and communicative processing (such as the higher association cortex) until after the first year infancy, by which time pruning will have already commenced in low-level sensory and motor areas (Huttenlocher & Dabholkar, 1997; Huttenlocher, 2002). This could explain why declines in social and communicative skills have not been concretely identified in high-risk studies and why atypicalities are identified across multiple cognitive domains by 14 months.

Models predicting 36-month diagnostic outcomes were also run in order to compare the stability of predictors of outcome over time. At 36 months, the one significant predictor of outcome was AOSI total score. That is, a higher AOSI score increased the likelihood of being categorised as having ASD in the high-risk category. Whilst it is likely that interaction effects between other parameters are also affecting the predictive value of the model, AOSI was the strongest single predictor at 36 months. As AOSI is designed to recognise and test for behaviours indicative of ASD, one possibility is that by 36 months, individuals with ASD are showing a greater range of autism-specific behaviours, whereas at 24 months, individuals could be displaying more general atypical behaviours.

Applying Phase 2 data to the logistic regression model validated predictive findings. When binary outcomes were generated and compared with 24-month ADOS classifications from Phase 2, individuals with the most severe ADOS scores, that is, those above the threshold for Autism (and not ASD) were more accurately classified, with a positive prediction value of 70.5% as opposed to 55% when those above the threshold for ASD were included. This

suggests that a subset of infants who are diagnosed with more severe ASD behaviours at 24 months could be identified through earlier atypicalities in comparison with infants who have a milder form of ASD. However, in order to elucidate this finding further, utilising data from a larger number of infants would be necessary. It will also be useful to compare predictive results at 36 months using Phase 2 data, when full diagnostic classifications have been completed.

Whilst the model was reasonably successful in predicting outcome, between 30% and 45% of individual outcomes were not classified by the model. There are several potential reasons for this. This number could, in part highlight the heterogeneity present in individuals with ASD. The variability in both the behavioural manifestation and aetiology of ASD makes the identification of predictive variables of models of developmental patterns highly challenging.

Alternatively, the sample size could explain why there was poorer predictive power in the Phase 2 validation models. When logistic regression models are applied to data, it is preferable to have an equal number of 'events' (in this case being classified as above the threshold for ASD) and 'non-events' (classified as below the threshold for ASD) in both phase 1 and phase 2 datasets. In phase 2 data, there are twice as many infants than are in Phase 1, but fewer infants with ASD overall (35% in Phase 1 versus 10% in Phase 2). Logistic regression relies on logit coefficients, which are prone to bias in smaller samples. Therefore, this could be causing 'non-events' (below the threshold for ASD classification) to overwhelm the data (King & Zeng, 2001). I plan to determine the extent of this potential bias by identifying a 'rare event' correction that could be applied to the intercept of the regression model in order to account for the differences between the datasets.

Due to the modest sample size, and the fact that Phase 2 data were only available at 24 months, I consider these findings preliminary. Secondly, as all infants were from at-risk families, that is, families in which at least one other family member has an ASD diagnosis) is difficult to generalise predictions. For future iterations of the model, it will be important to create a more fine-grained environmental predictor of development. Lower SES levels have been linked to poorer developmental outcomes (Bradley & Corwyn, 2004) and higher SES levels associated with significantly higher recovery levels after intervention training in ASD (Carr et al., 2015). In the current study, environmental factor 1 was related to the successful prediction of developmental outcome. However, in the BASIS dataset, full demographic information was not provided for all infants and a small number of measures were collected overall. Phase 2 included a larger number of environmental measures, which I would use to assess the relationship with outcome and other variables. One statistical limitation was the abundance of ANOVAs used to identify the most appropriate measures to include in the regression models. Separate analyses were conducted for all Mullen and Vineland measure, which could have increased the occurrence of, type I errors. Whilst significance values were adjusted in each ANOVA to control for such errors, more parsimonious models could be constructed to minimise the multiplicities in future analyses.

This aim of this chapter was to use multiple risk factors to predict the occurrence of ASD in a high-risk population. This was one of the first high-risk studies to provide evidence for multiple-factor models of ASD. This is supported by empirical findings from high-risk studies, demonstrating atypicalities across domains. Despite limitations, the current findings support the consensus from prospective research, that few differences have been identified in cognitive domains before 12 months. Here, the earliest differences were identified at 14 months and were widespread across social and non-social domains. When added to a single

regression model, these behavioural distinctions were predictive of later diagnostic outcomes. Overall, these findings provide evidence for a predictive model of outcome in a subset of infants with ASD. Specifically, the current results provide preliminary evidence that atypical behaviour in motor and expressive language domains combined with lower scores in SES measures was predictive of outcome at 24 and 36 months. Whilst expressive language and fine motor both independently predicted later ASD classifications, a cumulative model that incorporated both significant predictors and AOSI and environmental factor 1 was a stronger predictive model. AOSI and environmental factor 1, whilst not independently significant, did improve the predictive ability of the model, suggesting the presence of higher-order interactions among predictors. Overall the model demonstrated that a combination of predictors had a cumulative impact on risk. Furthermore, the validation from a second independent dataset suggests that using models such as this in future may be useful in elucidating the underlying interplay and predictive power of cognitive and environmental processes in infants at risk of developing ASD. Understanding how risk factors combine in the development of the manifestation of ASD can inform targeted intervention strategies and therefore have implications for later prognosis.

Chapter 4

4.1 Introduction

As briefly discussed in Chapter 1, a number of neurocomputational models have been proposed with the aim of elucidating the mechanisms underlying behavioural deficits in ASD. In this chapter, I provide an overview of these computational models and discuss the importance of utilising a developmental computational framework to understand the influences on developmental trajectories and the identification of homogenous subgroups in ASD. I extend findings of the computational Over-Pruning account (Thomas, Knowland & Karmiloff-Smith, 2011; Thomas Davis, Karmiloff-Smith, Knowland & Charman, 2015) as described in chapter 1 to investigate whether homogenous subgroups could be explained by a single pathological mechanism, over-pruning, interacting with individual difference factors.

I create two novel populations and utilise one original population from the Over-Pruning account framework to identify regressive and non-regressive subgroups based on developmental trajectories. I then utilise statistical regression models to identify the computational parameters that account for the between-group variability. I then explore recovery rates in regressive individuals and identify the parameters that influence recovery. The neurocomputational properties underlying good and poor outcome scores in the regressive and non-regressive simulations are also assessed.

4.2 Computational models of ASD

Artificial neural networks

Many of the models we consider here employed artificial neural networks. Artificial neural network (ANN) models are statistical learning models, which attempt to simulate

development in single cognitive domains. ANNs are a computational approach motivated by neuroscientific findings based on the human nervous system, and loosely modelled on the principles of neural information processing. The majority of computational models are not intended to demonstrate neural functioning, but are instead used as cognitive models that embody some of the key principles of processing such as excitation and inhibition (Mareschal & Thomas, 2007). Computational models can greatly benefit developmental psychology and complement experimental findings. One advantage of using ANNs to model development and developmental disorders is that they acquire their represented states through a learning process when exposed to a standardised /studied training environment. Furthermore, the learning process is influenced by network parameters and the nature of the training environment. Computational models present the possibility to explore a range of populations and behaviours that may not be possible in clinical populations. Identical populations can be used in multiple training environments to compare performance levels across conditions. It is also possible to identify the mechanistic causes of distinct developmental trajectories across multiple simulated populations. Unlike empirical findings, all variance can be accounted for by the neurocomputational parameters in a computational model.

Simulations are comprised of networks of interconnected processing units (also known as nodes) and the process of learning can be compared with the strengthening of synaptic connections through the communication of neurons (Thomas, Ronald & Forrester, 2011). The processing units are connected in layers and information travels through an input layer, one, or several interconnected layers and an output layer. Each connection between layers has a weight value that is initially randomly assigned. As activation passes through the networks, they undergo a process of training by repeated presentation of input-output pairs, whereby stimuli are presented a number of times, at each of which, the model's weights are adjusted

incrementally. Disparity between activated output and target output is used to alter the weights to bring the actual output closer to the target. This is carried out for all items in the training set, then repeated to find the compromise weight matrix for all input-output pairs.

ANNs typically contain some form of 'learning rule' used to train the networks, specifying how weights are changed in response to inputs. In this thesis, the backpropagation of error algorithm (Rumelhart, Hinton & Williams, 1986) was used and is described here. For a network, the aim of training is to learn a single set of weights so that any input pattern is able to produce the correct output pattern. It is also desirable that these weights will allow a network to generalise to novel data that was not encountered in the training set. The weights of the network are initially randomly assigned, and so the outputs generated at the beginning of training are unlikely to correctly match those that go with the input pattern. As the neural network is 'guessing' what the pattern may be, initially, outputs are likely to be random. The role of back propagation is to make appropriate adjustments to the connection weights after comparing the network's desired output with the network's actual output. Each node in the network is compared in this way in order to calculate the network error with the aim of changing the weight connections to reduce error on this pattern (Plunkett & Elman, 1997). Errors for internal nodes are derived from an algorithm that computes their degree of blame for errors on the units in higher layers to which they are connected. Initial changes are small in order to be able to find a set of weights that are able to account for all patterns rather than the only correcting for a specific pattern. This procedure continues through the network until the layer above the input layer is reached. This process is a type of supervised learning and occurs for the duration of training.

4.3 Computational models of ASD

A number of computational models have been proposed to identify some of the mechanistic processes in ASD. Below, I provide an overview of salient existing computational models and consider the advantages of the Over-Pruning account of ASD (for other related models, see Dovgopoly & Mercado III (2013); Church et al., (2010); Gustafsson & Paplin´ski (2004); Andrew & Gustafsson (2003); Revithis & Tagalakis (2012); De Carvalho et al. (1999); Noriega (2008); Kriete & Noelle (2015); Chonnaparamutt & Barakova (2008); Duch et al., (2012, 2013)).

The majority of computational models exploring ASD have focused on specific deficits attributed to the disorder, rather than overall mechanistic differences and cascading effects that could be account for the full cognitive profile of ASD and the broader autism phenotype. Such deficits include atypical category formation, memory function and over-detailed processing of features.

One category of models has identified differences in feature recognition, and specifically, atypicalities in neural codes that are utilised in representing concepts such as semantic memory. One such model is that of McClelland (2000), which explored the concept of hyper-specificity in ASD using neural networks. It has been argued that in general, those with ASD have the tendency to demonstrate extreme specificity when understanding novel concepts and representations (see Chapter 1). The model explains this rigidity with evidence based on the neurocomputational properties of neural nets and in particular, generalisation. This refers to the use of pattern overlap in order to build representations of associations when common elements of a specific input are experienced. The model also proposes a second mechanism, known as conjunctive coding. This reduces interference from overlapping pattern activations that incorrectly over-generalise stimuli that should be viewed as distinct

items. When conjunctive coding is introduced, individual networks only respond to combinations of firing patterns where a number of rules are met, as opposed to a general response to common elements. McClelland (2000) suggests that in cases of typical development, appropriate application of both generalisation and conjunctive representations allows individuals to extend what they have learnt to novel and similar experiences. In cases of ASD however, specific brain areas may use excessive amounts of conjunctive neural coding, and are less able to detect overlap between new inputs. It is therefore more likely that distinct representations would be allocated to novel inputs. This ultimately leads to an inability to generalise and utilise prior experiences.

A related computational theory also focused on atypical neural coding and context utilisation (Beversdorf, Narayanan, Hillier & Hughes, 2007). The model, based around simplified principles of the McClelland (2000) model of ASD, similarly explored the inability to generalise from previous experience and utilise contexts appropriately, due to restricted semantic associative networks and therefore sparse connectivity levels. The authors sought to capture the results of a prior empirical study. Beversdorf et al. (2007) compared results from ASD and typically developing individuals using the false memory task (Roediger & McDermott, 1995). Individuals were given a list of words, all of which were semantically associated. When recalling the presented words, critical lure items (that is, words that did not appear in the original list but were semantically associated) were also presented. Typically developing individuals were likely to create a false memory when viewing the critical lure words, showing susceptibility to the influence of associated items.

ASD groups exhibited decreased susceptibility to the influence of semantically related words and were therefore less likely to experience false memories. It was argued that the decreased

and impaired utilisation of context demonstrated in those with ASD allowed for superior performance on the task. The model utilised the semantic memory data and results matched the experimental findings. When a word was presented, semantically associated network nodes were activated less frequently in the ASD group, suggesting a decreased use of context and therefore generalisation. While the concept for the model is interesting, the results are reasonably limited due to the simplistic methods employed. For example, a lack of hidden units meant that learning could not be built into the model. Therefore, the use of a more complex learning system would help to further clarify such results.

A third hypothesis related to the theory of atypical generalisation is motivated by evidence from neuroanatomical studies of individuals with ASD. A number of hypotheses have proposed atypical brain connectivity as a causal mechanism in ASD (Abrahams & Geschwind, 2008; Belmonte, 2004). Whilst some of these theories have suggested a reduction in functional connectivity in individuals with ASD (Just, Keller, Malave, kana & Varma, 2012; Koshino et al., 2008), others have pointed to evidence of local over–connectivity (Courchesne & Pierce, 2005; Vissers, Cohen & Geurts, 2012). Cohen (1994, 1998) used a back propagation technique to elucidate the effect of having too many or too few internal connections. An abundance of simulated connections led to poor generalisation and an inability to discriminate between groups. Similarly, in the model by McClelland (2000), an abundance of connections meant that discrimination ability was very good. However, generalisation ability was inferior due to over-detailed representations. Similarly, Grossberg and Seidman (2006) used what they called an iSTART model, which attempted to clarify how autism could arise from imbalances in brain mechanisms that control motor, attention, and cognitive behaviours, ultimately leads to the over-allotment of resources in representing categories.

Gustafsson (1997) discussed the use of self-organising feature maps for future computational models of ASD, which he argued could be used to simulate cortical processing. Comparing the columnar structure of the cortex to the structure of feature maps, Gustafsson proposed that feature maps in ASD might be narrow, ultimately leading to difficulties detecting features in stimuli unless there is very little variability. Empirical data from central coherence theories, difficulty with facial recognition and uneven cognitive profiles were accounted for with the suggestion of differing levels of impairment across different cortical feature maps.

Lewis and Elman (2008) used a back propagation ANN to suggest that deviations from typical brain growth trajectories would ultimately lead to atypicalities with long and short-range brain connectivity. Head circumference charts of typically developing children and children with ASD were used in a network that simulated the growth of inter-hemispheric interactions. By measuring the growth of such interactions at 4 time points between 12 and 48 months, the autistic models demonstrated brain reorganisation that was driven by brain overgrowth. The impact of this was disruption to typical patterns of connectivity and behavioural impairments in the trained model.

More recently Pellicano and Burr (2012) have proposed an account to explain the non-social and sensory symptoms of ASD using Bayesian decision theory, a statistical approach that quantifies trade-offs between classification decisions using probability and the costs that accompany those decisions. It is proposed that people with autism do not demonstrate atypical sensory processing, rather they do not employ the typical mediation of sensory input by internalised expectations of the world. In Bayesian theory, this is known as a prior. This prior is a probability distribution that represents the uncertainty about a parameter before incoming information is taken into account. The authors propose that autistic perception

results from atypicalities in either the construction of the prior probability or at the point where priors are combined with incoming sensory information, which would create greatly reduced or hypo-priors. Hypo-priors would ultimately lead to a reliance on incoming sensory information (bottom-up processing), causing autistic individuals to perceive the world in a more 'accurate' way, as they are influenced less by the bias of prior experiences. This has been suggested as a potential explanation of hypersensitivity, for example. The authors also argue the account could be used to identify testable hypothesis regarding atypical perception in ASD using research from visual perception tests and adaptation studies to support their account.

The aberrant precision model (Lawson, Rees & Friston, 2014) followed the Bayesian model by Pellicano and Burr (2012). Lawson and colleagues proposed that there might be difficulties in attenuating sensory precision and contextualising sensory information in ASD. The authors propose that atypicalities in the perception of information, and dynamics of social behaviours and actions can be explained by an imbalance of precision attributed to sensory information that has stemmed from prior beliefs. Precision is defined as a mechanism that controls the influence of prior beliefs in relation to sensory information. High sensory precision increases the influence of information from sensory channels that we consider to be reliable, whereas a low sensory precision shows a bias towards prior beliefs. They attempt to link sensory difficulties to problems in social and communicative domains with the hypothesis that the greatest difficulties will emerge for individuals with ASD when environmental uncertainty is high, for example during social exchanges. The authors suggested that atypical neurobiological mechanisms, such as hormone and neurotransmitter function, could be a part of the underlying cause of a difficulty in precision. The role of oxytocin, for example, was suggested to result in atypical attenuation of multimodal cues if it is functioning atypically.

The current thesis focuses on the Over-Pruning account of ASD (see chapter 1 for a full overview). The model originated from an artificial neural network model using population-modelling methods across large (N=1000) populations of simulated individuals, whereby the learning ability of individual networks was tracked across longitudinal trajectories. The original model proposed that a single pathological mechanism interacting with population-wide variation could produce multiple subtypes of autism. Further, it suggested that risk and protective factors could be inherited without the pathological pruning mechanism, explaining the appearance of some symptoms without the full disorder, as seen in the BAP.

The computational models of ASD described in this section have provided a relatively broad range of theories based on numerous behaviours identified. However, none have convincingly explained the full range of variability seen across individuals with ASD, or provide an explanation that could be generalised across the full spectrum, in particular, those in the broader autism phenotype (BAP) who share some behavioural and cognitive difficulties. Moreover, the many of the computational models of ASD fail to implement a developmental process, or consider how this might be altered in ASD. Many of the computational models suggest atypicalities in neurobiological mechanisms identified in adults with ASD as a starting point for understanding deficits. None of the models above acknowledge the importance of brain plasticity in infancy, or the complex interplay of environmental, social, and biological interactions in the years of life. In order to implement a developmental process, data must be modelled longitudinally and networks must be measured and compared across multiple time points in development in order to capture the emergence of dynamic interactions that alter across development (Karmiloff-Smith, 2009). It is through understanding this process that underlying causal mechanisms can be identified. My work can address this by utilising population modelling techniques to capture developmental processes in multiple computational populations.

Here, the goal of these simulations is to compare the high-risk population from the first iteration of the Over-Pruning account with two novel populations, one with an earlier onset on pruning and the other a control population. Using the early onset population, subgroups are identified by looking at individual developmental trajectories, and the underlying neurocomputational parameters specific to each subgroup are assessed. I also investigate recovery post-pruning in regressive simulations.

4.4 The Over-Pruning account of ASD

Previous findings

This chapter aims to identify potential mechanisms underlying variability in ASD by extending the original Over-Pruning model (Thomas et al., 2011) to examine other non-regressive developmental trajectories in ASD. The 2011 model focused on regressive subtypes in ASD (whereby regression was defined as a drop in accuracy levels after a stage of typical development) and demonstrated that a single neurocomputational parameter was able to produce developmental regression when set to atypical severity levels. This was the pruning threshold parameter, which determined the strength of a parameter before it deemed as unused (and was therefore available for pruning). One hypothesis from these findings was that the same model could explain other subtypes of ASD if the pathological pruning mechanism interacted with individual differences in other parameters, such as the onset of pruning. However, this was not demonstrated in the model. Whilst the model identified regressive networks, it did not clarify whether non-regressive yet atypical subtypes existed within the population, nor did it investigate recovery from regression or long-term outcomes. In the current chapter, new populations were created and analysed to address these questions.

4.5 Novel populations

A total of three populations are utilised in the current chapter; the first, a high-risk population (HR) originates from the 2011 paper, in which individual networks were at higher risk of regression overall. Additionally, two novel populations were created for the purpose of this thesis. The first of these was a high-risk, early onset population (HREO) whereby individuals were at a higher risk of over-pruning and were more likely to have an earlier onset of this pruning mechanism. This population allowed for the evaluation of whether differential timing of onset of over-pruning could produce non-regressive subtypes of ASD, for example, 'early onset' or 'late onset' developmental patterns (Landa & Garrett-Mayer, 2007; Elsabbagh & Johnson, 2010). The second population was a high-risk early onset control group (HREOC); identical to the HREO population apart from aggressive pruning was set to a low typical value. This lowered the risk of severe pruning, but maintained an early onset of pruning. One aim of this chapter is to understand the role of timing and regression the HREO population. The HREOC was selected as the optimal control population to compare with the HREO networks. All networks were identical in both populations apart from the pruning onset parameter and (which was set to typical values in the HREOC population) and the pruning probability. A unique advantage of modelling is that precise control populations can be created in order to compare how an individual would have developed in the same environment with all other parameters the same but for the case of the pathology. This allows for specific consideration of the role of pathology in the face of other individual differences factors.

4.6 Research questions

(1) Subgroups

Previous findings in the Over-Pruning account have proposed but not demonstrated that other subgroups of ASD could potentially be explained by a single pathological mechanism

interacting with individual differences. Three populations are analysed here to evaluate these claims. The aims are as follows:

(i) To validate the HREO population against the HR population. As both populations share the same parameters apart from onset of pruning, this can inform us of the impact of the timing of pruning.

(ii) To identify non-regressive subgroups based on developmental trajectories in the HREO populations, and to implement statistical models to identify the parameters that differentiate each of the groups.

(iii) To assess the extent to which HREO subgroups are differentiated by other individual factors.

(2) Outcomes

(i) For atypically developing networks exhibiting regression: To evaluate acute recovery from over-pruning induced deficits by grouping simulations by recovery rate, and using statistical models to identify the parameters predicting different rates of recovery.

(ii) To assess long-term outcomes from over-pruning induced deficits by comparing individuals with high and low outcome scores, and to identify the parameters that predict outcome.

(iii) For atypically developing networks not exhibiting regression: To assess long-term outcomes for non-regressive simulations by comparing individuals with high and low outcome scores, and to identify the parameters that predict better or worse outcomes.

4.7 Methods

4.7.1 Design and properties

A population modelling approach was used in the simulations implemented in this thesis (Thomas, Baughman, Karaminis, & Addyman, 2012), with the aim of capturing the development of large populations of individuals. The methods employed in this chapter were implemented from a training set used in the work of Thomas, Knowland & Karmiloff-Smith (2011). Development was simulated in 1000 artificial neural networks in each population, and included variation produced according to the quality of the learning environment and an individual network's learning properties. The learning properties consisted of 14 neurocomputational parameters that could each alter the capacity or plasticity of the learning networks. This, along with a manipulation of the learning environment, created individual differences in behaviour. The interaction of these parameters determined the specific learning abilities of each individual. Each parameter related to how the network was constructed, activated or adapted. Each network was run for a total of 1000 epochs; (where, one epoch was an exposure to all patterns in an individual training set), at which point a final outcome score was recorded.

4.7.2 The learning problem and training set

The simulations employed connectionist pattern associator networks that were trained using a supervised backpropagation-learning algorithm. This population modelling technique was used to simulate the acquisition of the English past tense formation. The past tense training set was considered only as an abstract mapping problem within cognition. This type of architecture has been used in a number of cognitive models of development (see Thomas & Mareschal, 2007). The mapping problem was quasi regular, in that it included a predominant

regularity, which could be generalised to novel input patterns along with a set of exception patterns. The learning environment enabled the assessment of the role of similarity and type frequency in development. Through these properties, the domain was taken to be representative of some mapping problems that could arise in a cognitive system, including language formation and category formation. The training set contained 508 training patters and was defined over 90 input units and 100 output units using binary coded representations. A generalisation set comprised of 410 patterns was also included in the training.

The training set was an artificial language designed by Plunkett and Marchman (1991) to appropriate properties of the English past tense problem. It comprised monosyllabic constructed around consonant vowel templates and encoded in terms of articulatory, phonological features. There were five types of verb stems in the training set: the regular mapping required a network to reproduce the input pattern on the first 90 units of the output layer and add a binary code on the final 10 units of the output layer. A total of 410 regular patterns were included in the training set. The regular pattern had a high type frequency and was referred to as ***regular.*** This type of verb formed the past tense by adding allomorphs of the +ed rule (e.g. watch-watched, chat-chatted). In order to test the generalisation abilities of each individual network, 410 novel patterns were included in the training set. These shared 60 of the 90 input elements with the ***regular*** patterns. This set was referred to as ***rule.*** These verbs focused on the generalisation to the regular rule, but to novel stems (e.g. wug-wugged). Three exception mappings were included in the training set. The three classes of patterns were increasingly dissimilar to the ***regular*** mappings. The first exception class was the most similar to the ***regular*** mappings, however all three classes had a lower frequency type. The low frequency type created a level of difficulty, and so the three classes were referred to as ***hard, harder and hardest***: to highlight the gradient of task difficulty. The first reproduced

the input but did not add the final code, and a total of 20 such patterns were included. This was referred to as ***hard.*** These verbs past tense form were identical to the verb stem (e.g. hit-hit). The second exception mapping reproduced a proportion of the input and did not add to the code. A total of 68 patterns were included, and was referred to as ***harder***. The past tense was formed in these verbs by changing an internal vowel (e.g. hide-hid). The third exception mapping associated an arbitrary binary pattern with the input and did not add the final code. A total of 10 patterns were included in the training set and was referred to as ***hardest.*** These verbs had no relation to the verb stem (e.g. go-went) and in order to be learnt they had to have a higher token frequency.

4.7.3 Neurocomputational Parameters (variations in learning capacity)

Variation across the simulated populations occurred because of differences in learning properties, and learning environment. These variations were produced through the interaction of 14 neurocomputational parameters, which were used to explore the identification of subgroups and variation in recovery across populations. In principle, the parameter settings allowed for 2,000 billion unique combinations. Table 4.1 provides a description of all parameters and their variation ranges.

Table 4.1 provides descriptions and the variation range for each of the 14 neurocomputational parameters in the Over-Pruning computational model of ASD.

	Definition	Variation range
Hidden Units	Hidden units modulated the computational power for each network. A larger number of hidden units within a layer increased the computational power and therefore increased the learning speed. However, more layers of hidden units would increase computational power but decrease the speed of learning, as errors would need to be spread from the output to deeper layers of the network in order to improve learning.	For networks with a hidden unit layer, the number of hidden units was varied. The minimum number of hidden units was 10 and the maximum was 500.
Architecture	The Architecture parameter determined the number of layers of connection weights that existed in an individual network. A 2-layer network does not have a layer of hidden units and can therefore learn more quickly, but has less computational power. A 3-layer network incorporates a layer of hidden units, providing greater computational power but a lower learning speed. A fully connected network is comprised of direct connections from input to output and a layer of hidden units, which produces a computationally more powerful system than both 2-layer and 3-layer networks.	Networks could have 1, 2, or 3 layers of connection weights and were encoded as 1,2 and 3. As with all parameters, a lookup table and random value generation was used to generate the value for each individual, with 'medium' values being more likely in the population.

Sparseness	The Sparseness parameter determined the probability that any given connection would be created in a single layer of connection weights. It therefore modulated the proportion of connections in each layer. A greater level of connectivity would lead to an increase in computational power.	The Sparseness parameter ranged between 50% and 100 % (of initial connectivity between layers).
Weight Variance	The Weight Variance parameter assigned the range of random connection weight values that each network was initiated with. If initial weights are larger, learning rate is slower due to the increased time it will take to these weights to unlearn.	This ranged from 0 to 4. For example, if the weight variance were set to 0.5, weights would be randomised between +/- 0.5.
Processing Noise	The total activation a receiving unit received from a single sending unit is a result of the connection strength between the units and the sending unit's activation. Gaussian transmission noise was added to this total activation in order to simulate the effects of many naturally occurring random processes. Elevated levels of noise have been suggested as a potential explanation for hypersensitivity and hyposensitivity in ASD (Simmons et al., 2007). Conversely, it has also been argued that ASD symptoms reflect too little noise, enhancing discrimination but at a significant cost (Davis & Plaistead-Grant, 2014).	Processing noise was implemented on a continuous scale from 0 to 6.

	A distinction is also made here between the processing noise parameter and 'noisy' individual trajectories that are seen in the 'Messy' category of developmental trajectories in this chapter. 'Messy' individual trajectories demonstrated fluctuations in performance, which was independent of processing noise.	
Momentum	Momentum allowed a proportion of the weight change from the previous learning trial to be carried over. This function is used to prevent learning from getting caught in sub-optimal learning solutions.	Momentum ranged from 0 to 0.75.
Nearest Neighbour Threshold	Correctness was established using the nearest neighbour threshold. The network output comprised of a vector of continuous activation between 0 and 1, and responses of the network were binary vectors. An algorithm determined which legal phoneme was closest to the activation patterns. However, the phoneme was recognised as a response only when the activation was sufficiently close to the legal phoneme using a root square measure (RSM). This was determined by the nearest neighbour threshold. A high threshold allows the approximate output to be accepted as correct (and a larger error is tolerated), whilst a lower	Nearest neighbour threshold values ranged from 0.0025 to 0.5.

	threshold requires a more exact initial output. The nearest neighbour threshold was used to convert continuously valued output actions into accuracy levels (also see Plaut et al., 1996 for a model of reading in which this method was derived).	
Connection Weight Decay	The size of each connection within a network was reduced at each presentation of a training pattern. The extent of this connection reduction was controlled by the connection weight decay parameter.	The range of the connection weight decay values were calculated by estimating a percentage of weight values that could plausibly be lost overall all of training (e.g. 50%), and dividing this proportion by the number of training epochs (e.g. 1000) and the number of training patterns presented on each epoch (e.g. 508), to provide a proportional reduction in the connection weights to be applied on each pattern presentation (e.g. 0.5/1000/508=9.84 x 10-7)
Learning Rate	Learning rate determined the extent to which the connection weights were altered in response to differences between the output and target during training. A high learning rate enables faster learning but at the expense of stability (performance) levels.	Learning rate values ranged between 0.005 and 0.5.

Learning Algorithm	The backpropagation algorithm (see page 118 for a description) or cross-entropy could be used to calculate the error signal (or disparity scores), which marked the difference between the network's current output and intended target. Cross-entropy was used as it was at a lesser risk of entrenchment.	
Pruning Onset	Connections that did not markedly influence network function were probabilistically pruned during training. The pruning onset determined the epoch at which pruning would commence. An epoch corresponded to the presentation of all the mappings in the training set of an individual network.	The pruning onset ranged between 0 and 1000 epochs.
Pruning Probability	Pruning probability determined the likelihood that a connection could be pruned if it was deemed not to be in use. Connections were deemed unused if their absolute magnitude fell below a certain threshold, as specified by the pruning probability parameter.	Pruning probability ranged between 0 and 1.
Pruning Threshold	The pruning threshold was the pathological mechanism in the model. The larger the pruning threshold, the more likely weights would be categorised as unused. In situations where the pruning threshold was set to extreme levels, useful circuitry could be damaged because connections were incorrectly identified as unused,	The typical range of variation for the Pruning threshold was 0.1 to 1.5. The atypical range of variation for the Pruning Threshold was 1.5 to 4.0. Individuals with Pruning threshold values in the typical range are defined as the HREOC population. For these

	therefore increasing the risk of over pruning.	networks, all values were set to 0.1 whereas individuals with pruning threshold values in the atypical range are assigned to HR populations where pruning is a pathological mechanism. Note parameter selection for each individual is probabilistic. For the HREO population, the range was between 0.1 and 4. Therefore, not all networks will have atypical values,
Temperature	A receiving unit sums the total activation from all sending units and uses an activation function to determine the output. In this chapter, a non-linear Sigmoid function was used. The function contains a parameter, the temperature, which makes the function steeper or shallower. A low temperature creates a shallow slope, which means that the processing units are less sensitive to small input differences. This ultimately leads to poorer discriminability and therefore slower learning, as units will have to rely on smaller differences in the input when learning to categorise. Conversely, if the temperature is high, a sigmoid will be similar to a step function and the majority of inputs will be either on or off ('saturated') rather than within their dynamic range. If a	The temperature parameter ranged from 0.0625 and 4.

	unit is saturated the sigmoid slope is flatter, which will lead to smaller weight changes for a given error signal and slower learning. A very high or very low temperature can therefore cause delays in learning.	
Environment/ Family Quotient	Varying the amount of information available for each network controlled the quality of the learning environment. Each simulation was assigned a family quotient, which was a number between 0 and 1 that related to the number of patterns in a training set that individual networks were exposed to. Scores were stochastically designated in order to represent the family or environmental condition that each network would be exposed to. A range was selected for each population and sampled randomly.	A higher family quotient score was indicative of a richer learning environment. For example, if an individual had a family quotient of .75, each training pattern that was presented had a 75% chance of being included in that individual networks' training set. A range between 0 and 1 would be used in an environmental-risk population. In the current populations, family quotients ranged between 0.6 and 1.

4.7.4 Parameter Variation

Two of the parameters were categorical: the architecture (the number of network layers) and the learning algorithm (backpropagation or cross-entropy). All other parameters were valued on a continuous scale. In order to create variation, sensitive ranges of each parameter were identified (Thomas, Forrester & Ronald, 2016). Each parameter on a continuous scale was varied in turn, whilst all other parameters were held at their initial values in order to identify a range of performance values. This resulted in a wide range of performances from failing to learn, to highly effective learning. The aim was to create a median value for each

neurocomputational parameter. Some of the parameters demonstrated the most successful learning as a value increased (e.g. more hidden units were better), whilst others produced optimal learning at an intermediate range (e.g. a mid-range temperature parameter provided a better rate of learning). Parameter values were more likely to be set at a value that produced medium performance levels rather than very good or very poor performance levels.

4.7.5 Pruning

Pruning was included in this model in order to capture an aspect of normal neural development, the elimination of brain connectivity occurring at different time points in different regions after early childhood. Pruning was not employed in other ANN models of ASD. Nevertheless, the model did not simulate the earlier elaboration of connectivity, only the outcome of this process (in the form of the sparseness parameter). Each individual's initial connectivity was specified. The model therefore implicitly assumes the elaborative phase of brain development to be normal in ASD. After a period of development (the timing was varied by the pruning onset parameter), pruning was initiated. Pruning was determined by three parameters, which varied between individual networks (Table 4.1 provides descriptions for each of the pruning parameters). Once pruning had commenced, each connection weight was measured against a threshold. If the connection weight fell below that threshold it was considered to be an unused or weak connection and removed with a probability quantified by the pruning rate. Once initiated, pruning remained active throughout each network's lifespan. However, a large number of individual networks demonstrated some degree of recovery following a phase of regression. The model was able to produce some form of recovery without removing the pathological process (the pruning threshold) as the remaining connectivity left in the network was used to continue learning using reduced resources.

4.7.6 Creation of populations

For each network, the 14-neurocomputational parameters were sampled from a distribution of each parameter (See Thomas, Forrester & Ronald, (2011) for further detail). Following this, networks were assigned a family quotient value in the range set for that particular population. The networks were then trained for 1000 epochs on its training set. Performance was assessed on all 5 mappings (***regular, rule, hard, harder, hardest***) at each epoch. Performance was measured using accuracy levels (percentage correct). In the current chapter, only ***regular*** mappings are analysed. One of the central aims of this chapter was identify non-regressive subgroups and the underlying mechanistic differences between groups. As this was an exploratory and initial analysis of a three large simulated populations, the focus was on the most common (***regular***) mapping to index the broad behaviours of each system.

Original population

In the original iteration of the Over-Pruning account (Thomas, Knowland & Karmiloff-Smith, 2011) three populations were created. The first was a population at low risk of regression (LR). This population had a reasonable learning environment (where family quotients were in the rage of .6 to 1) and the maximum pruning threshold was 1.5, with 4 being the overall maximum threshold. Randomised weight values that were assigned to each individual were sampled randomly from a Gaussian distribution of .5. The average pruning threshold for the low-risk population was 0.1, meaning that after a period of training; those with an average threshold of 0.5 were unlikely to experience a high volume of pruning. The second population, known as high-risk (HR), was at a higher risk of aggressive pruning and therefore, behavioural regression. For 13 of the 14-neurocomputational parameters, values were sampled from the same probability distributions as a low-risk control population. The

single difference in the HR population was derived from the pruning threshold parameter. That is, the probability of a higher pruning threshold value was greatly increased. The pruning thresholds for the majority of individuals in HR population were between .075 and 4, and therefore significant loss of connections was likely. Thomas, Knowland and Karmiloff-Smith (2011) argued that a later pruning onset served to strengthen connections before a high rate of pruning commenced and therefore serve to spare connections by strengthening the connection weights prior to the onset of pruning. A final population was created to investigate the impact of environmental risk (LR deprived). The same parameters were used as in the LR population with the exception of the family quotient range, which was between 0.6 and 1 in the LR populations, and between 0 and 1 in the LR deprived population. In the deprived population, individuals could potentially be exposed to a learning environment that was comprised of very few training patterns. This was one method of simulating an impoverished learning environment.

Current populations

In the original study, Thomas et al. (2011) focused on simulating behavioural regression. They argued that earlier onset of pruning might produce other apparent subsets of individuals. However, this was not demonstrated at a population level, nor did they explore the effects of differences in timing. In this chapter, further analyses were carried out in the HR population and two new populations were created in order to elucidate these questions. A high-risk early onset (HREO) and high-risk early onset control (HREOC) population were created to explore the interactions between pruning onset, severity and other individual difference factors. The neurocomputational parameters in the HREO population originated from the same probability distributions as the HR population apart from the pruning onset parameter. An increased likelihood of earlier pruning onset distinguished this group (see table 4.2). The HREOC population was identical to the HREO group in all respects except

for the pruning threshold, which was instead set to a low value (.1). A unique advantage of computational modelling is that it is possible to separate causes of variability. In the case of the Over-Pruning model, it is possible to identify whether variability stems from the pathological mechanism, individual differences, or a combination of both. The control population is ideal because it has identical parameter values to the HREO populations, however, it does not have the high-risk of pruning (the pathological mechanism). Therefore, comparing these populations enabled further understanding of the factors accounting for variability in cases of disorder. Table 4.2 shows the probability distribution for the pruning onset and pruning threshold parameters in the HR, HREO and HREOC populations (for the remaining distribution parameter values see appendix 3).

The HREO group exhibited a range of atypical trajectories, which provided an appropriate foundation to explore variability within regressive individuals and those showing atypicality. Furthermore, by analysing both the HR with the HREO group it was possible to determine the extent to which timing of pruning affected the overall trajectory, and whether a later pruning onset could potentially serve as a compensatory mechanism in some individual networks.

Table 4.2. *Probability distributions for generating pruning threshold and pruning onset parameters for individuals in the HR, HREO and HREOC populations. 1000 epochs indicates the end of training.*

Pruning threshold probability	**0.01**	**0.75**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
High-risk (%)	11	11.3	12.7	9.3	9.1	12.3	11.5	10.9	11.9
High-risk early onset (%)	10.6	10.6	12.2	9.6	13.7	11.6	8.9	10.5	12.3
High-risk early onset control (%)	100								

Onset probability (epoch)	0	20	25	50	75	100	150	250	500
High-risk (%)		0.08	4.6	10.4	17.3	47.2	12.4	6.2	1.1
High-risk early onset (%)	6	10.5	24.4	30.3	23.5	5.3			
High-risk early onset control (%)	6	10.5	24.4	30.3	23.5	5.3			

4.8 Results

4.8.1 (1) Subgroups

First, mean final scores were compared between the HREO, HREOC and the original HR population in order to explore differences between groups in terms of final level accuracy scores and the end of 1000 epochs of training. Where group differences did occur, statistical regression models were used to identify the parameters that predicted prognosis in each case. The same process was completed for non-regressive individuals. Three atypical groups and one typically developing group were identified based on each individual's developmental trajectories, and development of each group was examined using the same methods as the regressive group.

4.8.2 HREO vs. HR populations

Each of the 1000 simulated developmental trajectories in the HREO populated was coded to identify regressive and non-regressive individuals. Regression was defined as a drop in accuracy at any point in development *after* the onset of pruning. Regression could range from a temporary dip in performance levels to a permanent drop in performance levels. Individuals were categorised as a regressive individual when a drop in performance level was greater than the fluctuations present due to developmental variation. As in the original Over-Pruning paper (Thomas et al., 2011), trajectories were coded by hand, capturing the behavioural basis on which cases of atypical development are identified (i.e. blind to the internal properties of the

system. In a large proportion of the simulations, a higher level of internal processing noise caused oscillations between measurement points, and therefore automated regression coding was not possible to implement. A sample of 50% of the trajectories was rated by a second, independent coder. An interrater reliability analysis using the Cohen's k showed high agreement between the two judgements, k = .91 (95% CI, .844 to.943, $p < .005$). Several benchmark scales have been proposed to assess interrater Kappa values. Landis and Koch (1977) suggested a value between .81 and 1.00 indicates substantial agreement levels between raters. Everitt (1992) and Fleiss (1981) supported this benchmark and recommend that Kappa values between .75 and 1.00 indicate "excellent" levels of agreement. Kappa values above .80 were therefore were considered as demonstrating high reliability.

Other studies that have identified subgroups by analysing developmental trajectories have utilised subgrouping methods other than hand coding. One example is a growth trajectory modelling approach (Conti-Ramsden, St Clair, Pickles & Durkin, 2012) where statistical clustering models are run with increasing numbers of subgroups until the model no longer improves. In such models, the probability of how well an individual fits within a subgroup can also be measured.

Table 4.3 shows the number of individual networks that demonstrated regression or those that did not across the HR and HREO populations. The final score was the mean overall performance score at the end of training on the ***regular*** mapping domain (1000 epochs). Scores are provided for both regressive and non-regressive individuals. The number of individual networks that regressed was not significantly different between the HR and HREO group. Independent t-tests were used to compare final scores of the two populations from the ***regular***

pattern set. The HR non-regressive population showed significantly higher final scores than the HREO population $t(998)=2.472$, $p=.014$.

Table 4.3. The number of regressive and non-regressive individuals and final outcome scores for the HR and HREO populations.

	High-risk	High-risk early onset
Regressive (n)	641	602
Regressive final score (1000 epochs)	0.599	0.579
Non-regressive typical (n)	359	398
Non-regressive typical final score (1000 epochs)	0.855	0.81

HREO logistic regression analysis

A logistic regression analysis was used to elucidate the parameters that predicted the presence of regression in the HR and HREO populations (note, no cases of regression were identified in the HREOC population due to the low pruning threshold, so were not included in statistical regression analyses). The HREO population regression equation was statistically significant, χ^2 (14) = 673.68, $p < .005$. Of the 14 neurocomputational parameters, six were significant predictors of outcome. The results are shown in Table 4.4. An increase in hidden units, temperature, noise, pruning probability, pruning threshold and lower pruning onset predicted the likelihood of regression. Importantly, only elevated severity of pruning could be the direct cause of behavioural regression.

Table 4.4. Logistic regression parameters for the HREO group, predicting the outcome of regressive or non-regressive networks. Highlighted groups indicate significant predictors at the .05 level.

							95% C.I for EXP(B)	
Parameter	B	SE	Wald	df	p	odds ratio	lower	upper
Hidden units	**0.009**	**0.003**	**9.195**	**1**	**0.002**	**1.009**	**1.003**	**1.015**
Temperature	**2.055**	**0.264**	**60.385**	**1**	**<0.001**	**7.807**	**4.649**	**13.11**
Noise	**-0.489**	**0.114**	**18.313**	**1**	**<0.001**	**0.613**	**0.49**	**0.767**
Learning rate	-2.358	2.317	1.036	1	0.309	0.095	0.001	8.875
Momentum	-0.239	0.677	0.125	1	0.724	0.787	0.209	2.967
Weight variance	-1.192	0.251	22.614	1	0.055	0.304	0.186	0.496
Architecture	0.082	0.153	0.289	1	0.591	1.086	0.804	1.465
Learning algorithm	0.581	0.388	2.246	1	0.134	1.788	0.836	3.824
NN threshold	1.007	1.195	0.71	1	0.399	2.738	0.263	28.503
Pruning onset	**0.025**	**0.004**	**34.821**	**1**	**<0.001**	**.925**	**1.017**	**1.033**
Pruning probability	**3.082**	**0.575**	**28.728**	**1**	**<0.001**	**21.811**	**7.066**	**67.327**
Pruning threshold	**2.039**	**0.133**	**235.094**	**1**	**<0.001**	**7.685**	**5.921**	**9.974**
Weight decay	-0.962	1.629	0.348	1	0.555	0.382	0.016	9.308
Sparseness	-1.098	0.831	1.747	1	0.186	0.333	0.065	1.7
Environment	0.076	0.236	0.424	1	.0602	1.664	1.246	1.922

HR logistic regression analysis

In the HR population, the logistic regression analysis identified one significant predictor that distinguished the regressive and non-regressive populations χ^2 (14) = 151.01, $p < .005$. A higher pruning probability was associated with a greater incidence of regression. Results for individual parameters are shown in table 4.5.

Table 4.5 Logistic regression parameters for the HR population predicting the outcome of regressive or non- regressive individuals. Highlighted groups indicate significant predictors at the .05 level.

						95% C.I for EXP(B)		
Parameter	B	SE	Wald	df	p	odds ratio	lower	upper
Hidden units	0.022	0.013	2.866	1	0.090	1.022	0.997	1.049
Temperature	-0.414	0.397	1.088	1	0.297	0.661	0.304	1.439
Noise	-7.881	5.498	2.054	1	0.152	0.223	1.204	18.102
Learning rate	-2.288	1.824	1.573	1	0.210	0.101	0.003	3.625
Momentum	-0.535	0.622	0.741	1	0.389	0.585	0.173	1.981
Weight variance	-0.16	0.332	0.233	1	0.630	0.852	0.444	1.634
Architecture	-1.883	1.214	2.404	1	0.121	0.152	0.014	1.644
Learning algorithm	-2.489	2.939	0.717	1	0.397	0.083	0.012	26.338
NN threshold	0.454	0.004	0.001	1	0.977	1.023	0.993	1.007
Pruning onset	0.686	1.184	0.336	1	0.562	1.986	0.195	20.216
Pruning threshold	**3.183**	**0.474**	**45.053**	**1**	**<0.001**	**4.114**	**9.52**	**61.078**
Pruning probability	2.271	5.802	0.153	1	0.696	9.688	2.355	40.009
Weight decay	-2.476	2.207	1.259	1	0.262	0.084	0.001	6.354
Environment	-2.342	0.0214	1.074	1	0.078	0.324	0.004	0.481

4.8.3 Non-regressive subgroups and predictive parameters

A total of 365 individuals in the HREO population were identified as non-regressive. All non-regressive developmental trajectories were assessed individually. After viewing all trajectories, four distinct groups were identified and named Messy, Lower, Slower and Typical. Each trajectory was rated by eye and then compared to the criteria created for each subgroup. Messy trajectories demonstrated unstable noisy responses that oscillated continuously across development. Individuals categorised as having a Lower trajectory exhibited poor development over time, demonstrating accuracy levels of 40% or less at 500 epochs and a final accuracy level of under 60%. Individuals categorised as having a Slower trajectory exhibited slow development but higher levels of accuracy than the lower subgroup by the end of training, with final accuracy levels between 41% and 70%. Typical groups

demonstrated consistent development throughout training with final accuracy levels between 70 and 100%. To ensure consistency of group membership, a sample of 50% of the trajectories were blind-rated by a second coder. Interrater reliability scores were assessed using Cohen's k, which showed a high level of agreement, k = .90 (95% CI, .862 to.958, $p <$ .005. Figure 4.1 shows a prototypical example of each non-regressive subgroup trajectory that each individual network was compared with.

Figure 4.1 Prototypical examples of the four non-regressive subgroup trajectories. Each individual network was compared to these trajectories in order to assess group membership.

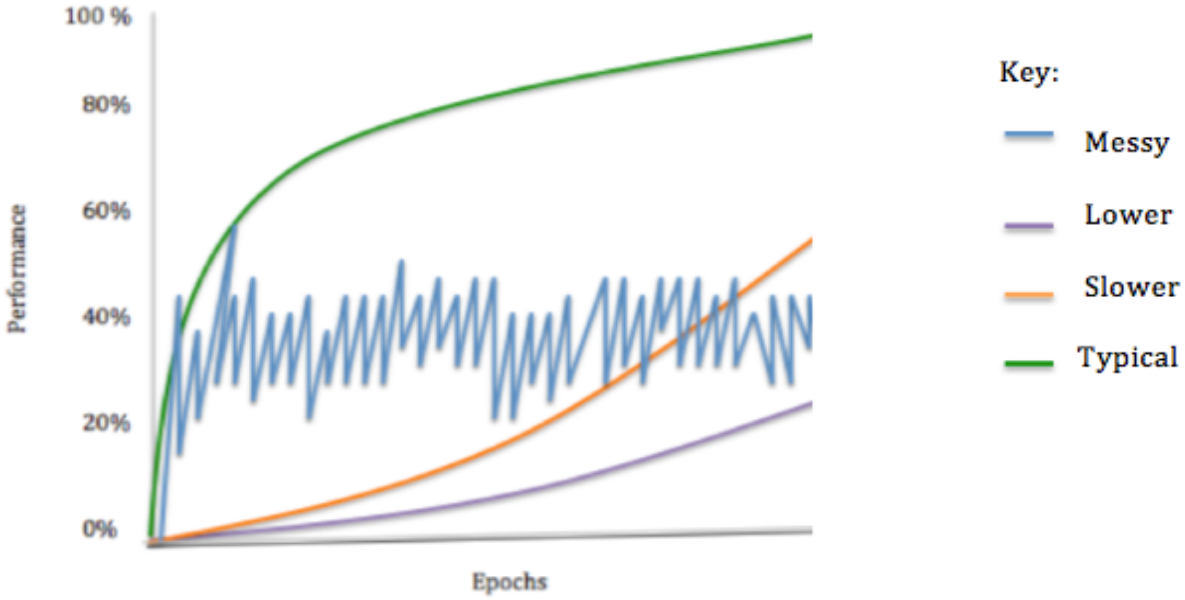
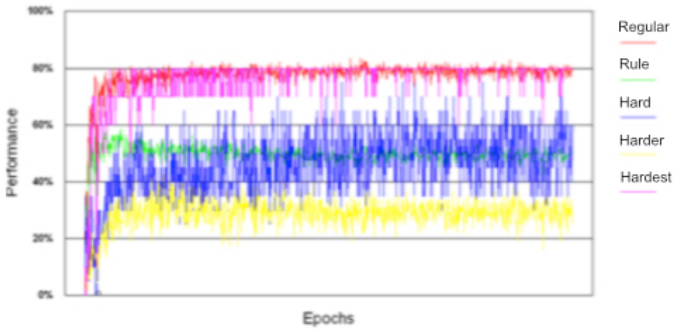


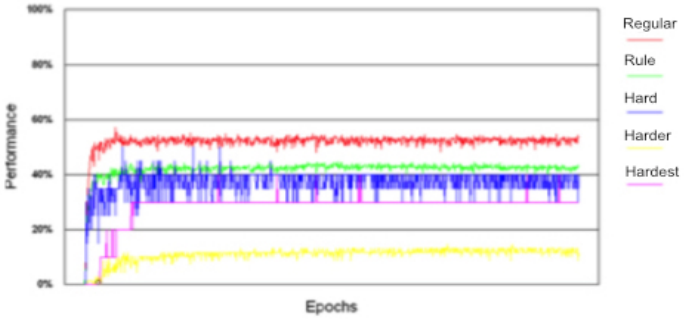
Figure 4.2 shows examples of individual trajectories for each of the four non-regressive sub groups across the five performance measures (***regular, rule, hard, harder and hardest***), although in these analyses the focus is on ***regular*** verbs only. A total of 31 networks were defined as messy, 34 were defined as lower, 90 were identified as slower, and 210 were identified as typical.

Figure 4.2. Individual example trajectories for each of the non-regressive subgroups: a) Messy trajectory b) Lower trajectory c) Slower trajectory d) Typical trajectory. Each trajectory profile shows development across the five measures for the duration of development (a total of 1000 epochs).

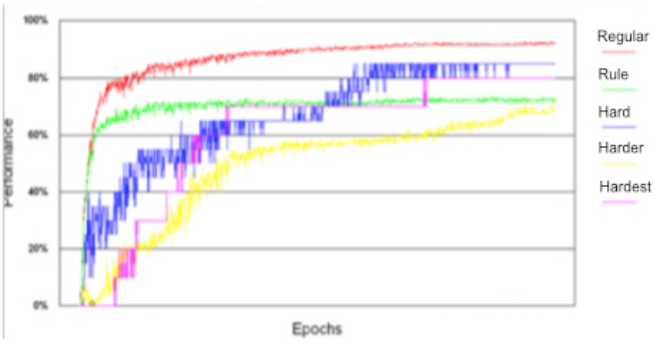
a)



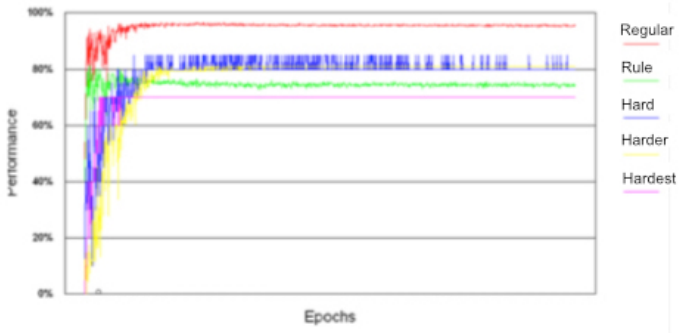
b)



c)



d)



4.8.4 Predicting group membership

In order to investigate the differences in development between subgroups at a mechanistic level, two types of analyses were run. In each case, I examined the extent to which each neurocomputational parameter affected group membership. First, a one-way multivariate

analysis of variance (MANOVA) was run to determine the effect of atypical grouping and to identify the effects of neurocomputational parameters between the four non-regressive subgroups (Messy, Lower, Slower and Typical). A total of 14 neurocomputational measures were assessed (see methods for a full explanation of the role of each parameter). Table 4.6 shows the mean neurocomputational values for the four subgroups. The dependent variable was subgroup type, and this had four levels. The independent variables were the complete set of mean parameter values. A MANOVA was the most appropriate option for these data, and it was therefore accepted that some loss of statistical power would occur. Levene's test indicated unequal variances *(*$F = 4.23$, $p = .015$); therefore, a lower level of significant α level of .005 was used.

Table 4.6. Mean neurocomputational parameters in the four atypicality non-regression groups. Parameters showing significant group effects are in bold (see text).

Parameter	Messy (n=31)	Lower (n=34)	Slower (n=90)	Typical (n=210)
Hidden units	60.97	62.21	59.17	71.31
Temperature	1.22	0.93	0.70	0.95
Noise	0.87	1.36	0.76	0.49
Learning rate	0.16	0.11	0.11	0.14
Momentum	0.30	0.28	0.23	0.28
Weight variance	0.66	0.61	0.69	0.62
Architecture	**0.58**	**0.47**	**0.48**	**0.15**
Learning algorithm	1.00	0.82	0.92	0.98
NN threshold	0.15	0.11	0.10	0.12
Pruning onset	30.32	34.56	39.00	43.43
Pruning probability	0.08	0.09	0.11	0.10
Pruning threshold	**1.18**	**1.47**	**1.10**	**0.74**
Weight decay	0.0 ($x10^{-5}$)	0.0 ($x10^{-5}$)	0.0 ($x10^{-5}$)	0.0 ($x10^{-5}$)
Sparseness	0.04	0.04	0.05	0.03
Environment	0.82	0.82	0.78	0.82

A statistically significant difference was found between the neurocomputational parameters affecting the four non-regressive subgroups $F(42, 873) = 1322.742$, $p <. 006$; Pillai's Trace .674; $\eta_p^2 = .225$. Pillai's Trace was used instead of Wilks Lamda due to the unequal sample sizes between groups.

Follow-up univariate ANOVAs showed that learning rate $F(3,302) = 11.487$, $p <.005$; $\eta_p^2 = .102$,architecture $F(3,302) = 13.167$, $p <.005$; $\eta_p^2 = .116$, pruning onset $F(3,302) = 9.552$, $p <.005$; $\eta_p^2 = .143$ and pruning threshold $F(3,302) = 16.637$, $p <.005$; $^2 = .127$) were significantly different between the four groups. A Bonferroni corrected α level of .005 was used in the analysis.

These results confirm the involvement of the pruning threshold in atypical trajectories that are nevertheless non-regressive; and that other population wide individual differences may contribute to the differentiation of typical and atypical trajectories. Notably the environmental parameter played no role.

Multinomial logistic regression model analyses

A multinomial logistic regression (MLR) was utilised as a second, complementary method to understand the interaction of parameters and group. The model demonstrated significant differences between all groups $14(350) = 266.9$, $p <.001$, Nagelkerke $R^2 = .665$. Likelihood ratio tests demonstrated the following significant predictors: hidden units ($p < 0.001$), temperature ($p < 0.001$), learning rate ($p < 0.001$), momentum ($p < 0.001$), architecture ($p < 0.001$), pruning onset ($p < 0.001$) and pruning threshold ($p < 0.001$). Table 4.6 provides the

mean values for each neurocomputational parameter from the four subgroups. Table 4.7a and Table 4.7b show the significant results from the MANOVA and MLR comparisons.

*Table 4.7a and table 4.7b show the neurocomputational parameters that significantly discriminated between non-regressive sub groups for the HREO population. Note. ANOVA= analysis of variance; MLR= multinomial logistic regression. Scores show comparisons yielding results with p<.05. * Indicates reliable effects at <.005. ** Indicates reliable effects at <.001.*

Table 4.7a

Parameter	**Messy vs. Lower**		**Messy vs. Slower**		**Messy vs. Typical**	
	ANOVA	**MLR**	**ANOVA**	**MLR**	**ANOVA**	**MLR**
Hidden units		**<0.001****				
Sparseness						
Weight variance						
Temperature		**<0.001****		**<0.001****		**0.002***
Noise						
Learning rate	**<0.001****		**<.001****		**0.003***	**0.032**
Pruning probability	**<0.001****					
Pruning threshold		**0.022***			**0.008**	**<0.001****
Architecture					**0.002***	
Learning algorithm						
NN threshold						
Momentum						
Pruning onset						
Environment						
Weight Decay						

Table 4.7b

Parameter	**Lower vs. Slower**		**Lower vs. Typical**		**Slower vs. Typical**	
	ANOVA	**MLR**	**ANOVA**	**MLR**	**ANOVA**	**MLR**
Hidden units						**<0.001****

Sparseness						
Weight variance						
Temperature						
Noise					**0.004***	
Learning rate			**0.002***	**<0.001***	**0.004***	**<0.001****
Pruning probability						
Pruning threshold	**<0.001****		**<0.001****	**<0.001****	**0.004***	**<0.001****
Architecture		**0.001***	**<0.001***	**0.001***	**<0.001****	
Learning algorithm						
NN threshold						
Momentum						
Pruning onset						
Environment						**0.001***
Weight Decay						

The neurocomputational parameters predicting atypical group membership varied across subgroups, particularly the distinction between the typical and slower groups. The largest effect sizes were learning rate, architecture and pruning threshold. Overall, the typical group exhibited a lower Learning Rate (producing slower but more stable learning) and lower pruning threshold, and were more likely to comprise of a 2- or 3-layer architecture (as opposed to fully connected). This enables the network to learn more quickly but provides less computational power).

Comparing typical and lower conditions, lower individuals tended to have fewer hidden units, a lower learning rate and higher pruning threshold. These parameters would suggest that a combination of slower learning and greater likelihood of significant connection loss was the cause of a lower atypical trajectory.

Comparing the messy and typical conditions, the messy groups showed a higher pruning threshold, higher learning rate and higher temperature. The messy subgroup differed from

the lower and slower groups on a smaller number of parameters. A higher learning rate led to unstable development in response to increased connections.

Comparing lower and slower conditions, the lower subgroups displayed a lower learning rate and temperature compared to the slower subgroup, while the strongest effect size for the slower subgroup was a lower learning rate. Finally, the group comparison demonstrating the fewest parameter differences was lower versus slower. A lower pruning threshold in the slower condition was the only parameter that predicted outcome group.

Overall, population-wide individual differences in parameters modulated the effects of the pathological pruning parameter to produce heterogeneous atypical profiles.

4.8.5 (2) Outcome

Currently, in reality, the mechanisms underlying development and recovery in regressive individuals are unknown. A number of studies have reported no differences in developmental outcomes between regressive and non-regressive individuals with ASD (Werner, Dawson, Munson & Osterling, 2005), where others have found that on average, infants who displayed signs of developmental regression were more likely to have a poorer outcome when compared to other non-regressive children (Kalb, Law, Landa & Law, 2010; Luyster et al., 2005). Thus far, however, the factors accounting for better outcomes in some individuals are unknown.

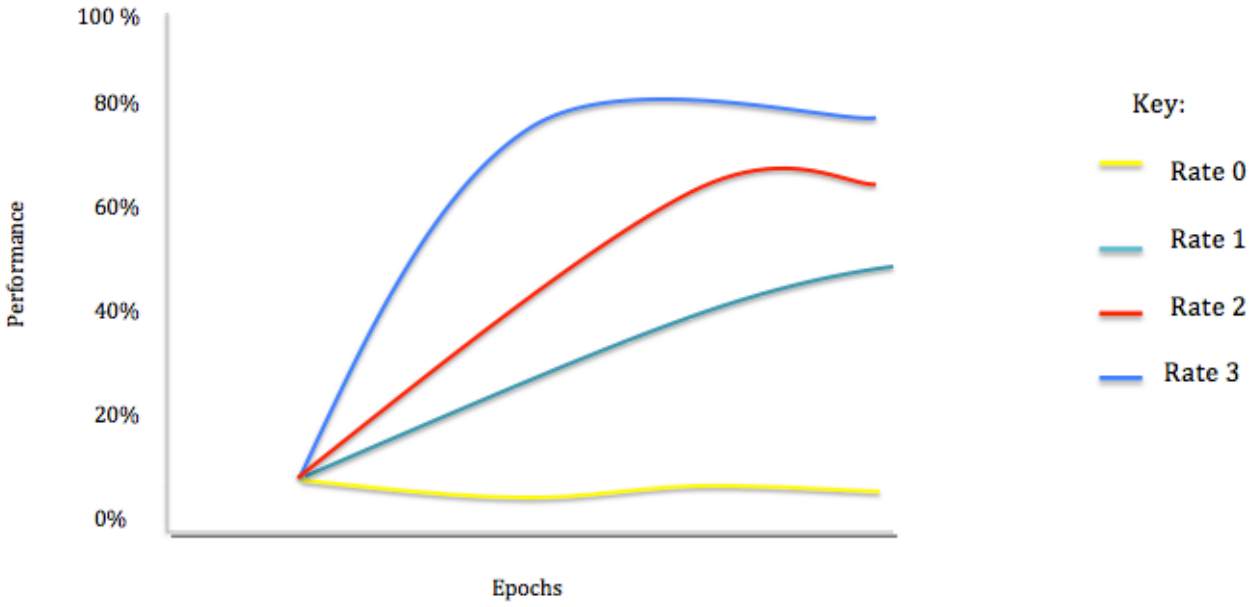
Here, regressive individuals were sub grouped based on the speed of recovery after regression had occurred. Multivariate analysis of variance and multinomial logistic regression models were used to identify the computational parameters were differentiating

the subgroups. High and low outcome scores were then compared to identify the mechanisms that accounted for better outcomes in regressive and non-regressive simulations.

4.8.6 Acute recovery from regression

In order to investigate the rate of recovery within the regressive individuals in the HREO population, each network was rated based on the slope of recovery observed in the ***regular*** measure after regression had occurred. A total of 602 cases of regression were identified, and four categories of recovery were defined: Flat recovery, (known as recovery rate 0) slow recovery (known as rate 1), medium recovery (known as rate 2), and fastest recovery (known as rate 3). Networks were compared to prototypical trajectories characterising each of the recovery rates. Figure 4.3 shows the prototypical recovery trajectories that were used to subgroup each individual. In order to confirm consistency within the group ratings, a sample of 50% of the trajectories were blind-rated by a second coder provided with the prototypes. Interrater reliability scores ranged from .84 to .88. When including all cases of regression in the HREO population, 28% of individuals were in the worst recovery group (rate 0), 16% were identified as rate 1, 39% were in the rate 2 recovery group, and 17% were in the highest group, rate 3.

Figure 4.3. Prototypical trajectories for the four subgroups determined by recovery rate following regression.



4.8.7 Predicting regressive recovery group membership

Here, the mechanistic distinctions between rates of recovery and how they are predicted by the neurocomputational parameters were examined. Table 4.8 shows the mean parameter values for the four recovery groups.

Table 4.8. Mean parameter values for the four recovery subgroups.

Parameter	Rate 0	Rate 1	Rate 2	Rate 3
Hidden units	68.768	65.986	66.299	81.928
Temperature	1.290	1.127	0.976	1.024
Noise	0.329	0.992	0.528	0.234
Learning rate	0.107	0.118	0.129	0.135
Momentum	0.234	0.218	0.262	0.266
Weight variance	0.595	0.606	0.495	0.566
Architecture	0.203	0.127	0.113	0.542
NN threshold	0.077	0.100	0.130	0.137

Pruning onset	44.384	47.887	50.343	53.494
Pruning probability	0.168	0.130	0.142	0.103
Pruning threshold	3.397	2.965	2.417	2.217
Sparseness	0.039	0.042	0.032	0.039
Environment	.79	.79	.81	.82

A one-way multivariate analysis of variance (MANOVA) was used to identify the influence of each of the neurocomputational parameters on the four recovery groups and elucidate the extent to which population-wide variation could be producing group membership. As with the atypical groupings, the full parameter set was included in all analyses.

A statistically significant difference was found between the neurocomputational parameters affecting the four recovery groups. $F(36,1449) = 1322.742$, $p < .006$; Pillai's Trace .674; $n_p^{2=}.254$. Follow-up univariate ANOVAs showed that Hidden units $F(3,540) = 3.96$, $p < .005$; $n_p^2 = .024$, temperature $F(3,540) = 13.482$, $p < .005$; $n_p^2 = .076$, learning rate $F(3,540) = 10.576$, $p < .005$; $n_p^2 = .061$, architecture $F(3,540) = 11.643$, $p < .005$; $n_p^2 = .066$, and pruning threshold $F(3,540) = 16.637$, $p < .005$; $n_p^2 = .229$ were significantly different between the four groups. The results of the MANOVA are shown below in Table 4.9a and Table 4.9b.

Table 4.9a and Table 4.9b also show the results from a MLR, used as a second complementary analysis for understanding underling parameter interactions between recovery groups. A significant difference between groups was shown $4(88) = 432.6$ $p < .001$, Nagelkerke $R^2 = .589$. Likelihood ratio tests demonstrated the following significant predictors: hidden units (0.048), temperature (0.014), noise (0.003), learning rate (0.002),

momentum (0.003), architecture (0.005), pruning probability (0.005) and pruning threshold (0.002).

Rate 1, 2 and 3 recovery groups differed significantly from the rate 0 individuals on a number of computational parameters. In all comparisons, the rate 0 subgroup demonstrated a higher temperature, lower learning rate, higher pruning probability and pruning threshold. This combination of parameters points to an intrinsic lower plasticity, and more severe pruning through a higher threshold and higher probability of removing connections rated as unused.

The fewest parameter differences were between rate 1 and 2 recovery conditions. The largest effect sizes were related to pruning threshold and learning rate. As expected, the rate 2 subgroups mean parameter scores were significantly lower pruning thresholds and a higher learning rate. Finally, the parameters contributing the differences between rate 3 and 2 and 1 were similar. Learning rate, pruning probability and pruning threshold showed the largest overall effect sizes across the three groups, which was indicative of a faster learning rate. The best acute recovery was a result of more hidden units and an architecture with more layers of connections, which increased the power of the network, a less severe pruning threshold and later pruning onset and lower pruning probability.

*Table 4.9a and Table 4.9b show the neurocomputational parameters that significantly discriminated between non-regressive sub groups for the HREO population. Note. ANOVA= analysis of variance; MLR= multinomial logistic regression. Scores show comparisons yielding results with p<.05. * Indicates reliable effects at <.005. ** Indicates reliable effects at <.001.*

Table 4.9a

Parameter	0 vs. 1		0 vs. 2		0 vs. 3	
	ANOVA	MLR	ANOVA	MLR	ANOVA	MLR
Hidden units					**0.048**	
Sparseness						
Weight variance						
Temperature		**<0.001****	**<0.001****	**<0.001****	**<0.001****	**<0.001****
			<0.000**		**<0.001****	
Noise		**<0.001****				
Learning rate		**0.003***	**<0.001****	**<0.001****	**<0.001****	**<0.001****
Pruning probability		**0.005**		**<0.001****		**<0.001****
Pruning threshold		**<0.001****		**<0.001****		**<0.001****
Architecture					**<0.001****	**<0.001****
					<.0001**	
Learning algorithm	**0.003***		**<0.001****		**<0.001****	
NN threshold						
Momentum						**<0.001****
Pruning onset						
Environment						
Weight Decay						

Table 4.9b

Parameter	1 vs. 2		1 vs. 3		2 vs. 3	
	ANOVA	MLR	ANOVA	MLR	ANOVA	MLR
Hidden units			**0.036**		**0.006**	**0.001****
Sparseness						
Weight variance						
Temperature		**<0.001****		**<0.001****		
Noise				**0.002***		**0.006**
Learning rate	**<0.001****		**0.05**	**<0.001****		**0.02**
Pruning probability				**0.002***		**0.049**
Pruning threshold	**<0.001****					**<0.001****
				<0.001**		
Architecture				**<0.001****		**<0.001****
Learning algorithm	**<0.001****			**<0.001****		
NN threshold						
Momentum	**<0.001****					**<0.001****
Pruning onset						
Environment						

Weight Decay	**0.03**	**0.005**

In sum, rate of recovery following an episode of regression was mainly explained by the loss of resources incurred, and the amount remaining.

4.8.8 Long-term outcome in regressive simulations

In order to explore the relationship between group membership and final outcome score, regressive individuals were split into two groups: high final outcome score (defined as the top 50%) and low final outcome score (defined as the lowest 50%). Groups were then compared using a MANOVA and multinomial regression. The independent variable was final outcome score, with two levels, and the dependent variables were the parameter values.

The MANOVA revealed a significant difference between the two lower final score groupings $F(36,1449) = 1322.742$, $p < .006$; Pillai's Trace .674; $n_p^2 = .254$. Follow up univariate ANOVAs showed that reliable differences in hidden units $F(3,560) = 2.646$, $p < .005$; $n_p^2 = .087$, temperature $F(3,560) = 13.482$, $p < .005$; $n_p^2 = .076$, learning rate $F(3, 560) = 8.64$, $p < .005$; $n_p^2 = .064$ and pruning onset $F(3, 560) = 4.610$, $p < .005$; $n_p = .142$).

A better long-term outcome was predicted by a lower pruning threshold and higher temperature and later pruning onset. It was also associated with a lower pruning threshold and a higher temperature and later pruning onset.

4.8.9 Outcome in non-regressive atypical simulations

Non-regressive individuals were split into two groups based on a mean split on final outcome score. Those with high final outcome score (top 50%) and low final outcome score (bottom 50%). Each low and high score group were then compared using a MANOVA and binomial regression. The independent variable was final outcome score with two levels, and the dependent variables were the parameter values.

A significant difference was found between the high and low final scores of the regressive networks $F(15,18) = 5.204$, $p < .005$; Pillai's Trace .739; $^2 = .583$. Looking at the individual parameters accounting for group differences, temperature $F(1,32) = 3.449$, $p = .003$; 1 $n_p^2 = .097$, learning algorithm, $F(1,32) = .941$, $p = .005$; $n_p^2 = .190$ and NN threshold $F(1,32)=.013$, $p = .005$; $n_p^2 = .012$ were significant predictors. A significant difference was also identified in the binomial logistic regression model 14 (588) = 266.9, $p < .001$, Nagelkerke $R^2 = .665$.

Overall, a better outcome was associated with a lower pruning threshold, higher temperature, learning rate and architecture, indicating importance of the causal pruning mechanism and a higher learning capacity. This differed with factors predicting long-term acute regressive subtypes in that recovery from regression was dependent on a later pruning onset, suggesting that networks would have a better chance at recovery if connection had more time to strengthen and develop before pruning. However, both positive regressive and non-regressive outcomes relied on a higher learning rate.

4.9 Discussion

This chapter explored subgroups and variability within simulated ASD populations under the assumption that the primary pathology is caused by over-pruning of network connectivity.

The effect of timing, specifically the onset of pruning was assessed by comparing rates of regression and outcome scores at the end of training across three populations. Networks were split into regressive and non-regressive groups, whereby regression was identified as a drop in accuracy levels in overt behaviour after pruning had been implemented. Regressive groups were further subtyped based on the recovery speed of individual networks and non-regressive groups were sub-grouped based on the shape of developmental trajectories. Both regressive and non-regressive groups were additionally split based on final overall scores to explore the parameters that determined high and low performance levels. Overall, distinctions between subgroups were identified, mechanistic distinctions were shown between the HR and HREO populations, and computational parameters that predicted final outcome HREO population were isolated.

Comparisons between the HREO, HREOC and original HR population showed significant differences in final outcome scores in both regressive and non-regressive atypical individuals. One factor that was explored was whether there would be a higher incidence of atypicality (defined by a non-typical but also non-regressive trajectory) in the HREO population in comparison with HR. Although the difference in the number of regressive individuals was non-significant, t-tests comparing overall mean difference in final scores suggested that regression was more severe in the early onset population, and also that regressive individuals showed a significantly poorer prognosis overall. Some clinical studies have shown that early regression is indicative of a poorer developmental outcome (Lord, Shulman & DiLavore, 2004) these simulated results are consistent with this claim within an implemented mechanistic framework.

Statistical logistic regression models were used to identify the mechanistic basis of regression in the three populations and to further investigate whether differences between the causes of regression could be leading to a better or worse final outcome. The HREO population also demonstrated a different combination of parameters that predicted regressive outcomes. An increase in learning rate and capacity (hidden units, temperature) combined with an earlier onset and higher probability of pruning (higher pruning threshold and probability) predicted regression. It is possible that even though learning rate was generally higher across the population, an earlier onset of pruning meant that connections had less time to learn and create stability among connections. While all three populations demonstrated regression, initial analyses suggested three different processes that predicted regression over atypicality.

Due to the range of variability in final levels and developmental trajectories in the HREO population, the dataset was utilised to examine development and subgroups within the non-regressive population. A total of three atypical and one typical subgroup were identified according to behavioural developmental trajectories and in each case the computational parameters that accounted for group membership were identified. The parameters that distinguished those in the typically developing group were a lower pruning threshold and a higher learning rate, suggesting that a low-level pathological process combined with a fast learning rate allowed individuals to develop stronger connections more rapidly. For all of the atypical groups, a lower learning rate was one of the strongest predictors of a worse developmental outcome. If a lower learning rate was inherited separately, we might expect to see delay in some families.

Furthermore, it can be questioned whether training beyond 1000 epochs would have led to stronger recovery? This is a question of whether (and when) trajectories plateau. In this

model, the final level of development was chosen by including a calibration phase in the model to measure the point at which development ceases. Performance levels were compared between 900 and 950 epochs and 900 and 1000 epochs during calibration. No differences in development were identified, suggesting that by this point development had plateaued. In the case of recovery, this would suggest that individual networks are not delayed, but instead exhibit qualitatively different developmental trajectories from networks displaying typical development.

Whilst a large number of individual networks demonstrated regression, some individuals achieved successful recovery. Thomas, Knowland and Karmiloff-Smith (2011) suggested that remaining connectivity within networks is utilised to aid recovery across the rest of development to the highest level possible (considering the vast reduction in resources). Four groups were identified based on the slope of recovery until the end of development: Rate 0 (flat), Rate 1 (slow), Rate 2 (medium) and Rate 3 (fastest) recovery. Statistical MANOVA and multinomial regression models identified differences between each group in terms of parameters and group membership. The rate 0 subgroup was defined by a poorer learning ability, a limited learning capacity and greater likelihood of aggressive pruning. With rates 1, 2 and 3, the results suggest that as learning rate and number of hidden units increase, and pruning threshold decreases, the more likely it is that a better recovery level will be achieved.

Both regressive and non-regressive individuals were separated into a high final score, or low final score group, and score was compared in order explore the variability within groups, and what defined the differences in final outcome. In the non-regressive groups, lower pruning probability, threshold, and temperature predicted higher scores, which was indicative of a higher learning rate. In regressive subgroups, a better outcome was associated with a higher

learning rate as indicated by temperature, a later onset of pruning was indicative of outcome. It seems that for the regressive subgroups, a later onset of pruning was more important for a better recovery, as connections would have had a greater chance of strengthening before the pruning commenced.

Subgroups were identified in both regressive and non-regressive populations, and the mechanistic properties underlying group differences were identified. Better outcomes were determined by learning capacity, rate of pruning and rate of learning. Whilst no differences were identified in the number of regressive and atypical individuals between HR and HREO populations, the most positive outcomes after recovery from regression or development in the atypical groups was in the HR population, demonstrating the negative effect of early onset pruning on the remainder of development.

Whilst the results are useful as a complementary method of understanding mechanistic distinctions in ASD, there are number of limitations. Firstly, the model is an abstract learning system using a single architecture and problem domain. Results from computational populations can be linked to behavioural data, however it is not possible to directly map neurocomputational parameters onto aspects of neural or cognitive functioning. For future pruning models it would be an advantage to consider methods that could make the model more realistic to the neural systems, for example, by allowing changes in the learning system to occur not just between parameters, but also the input and output representations in the model, and to have a closer link to generate empirical data for specific behavioural paradigms.

Second, for future iterations of the model, it will be important to utilise statistical group modelling techniques as alternative to the hand coding approach to identify trajectories. Whilst interrater reliability levels were high using hand coding, using statistical techniques where models run with increasing numbers of subgroups until the models reach optimal group levels will serve as a complementary analysis, and group memberships can be compared between techniques.

In this chapter, subgroups were identified based on the recovery rates from the ***regular*** mapping condition. A question for future analyses is whether we would see differential outcomes using ***harder*** (irregular) mappings, which are more sensitive to network capacity. In the original Over-Pruning paper (Thomas, Knowland and Karmiloff-Smith, 2011), recovery levels between ***regular*** and ***harder*** mappings were compared in the HR population. Recovery levels (when compared with pre-regression peaks) were higher for ***regular*** mappings compared with the ***harder*** condition (13.6% and 8.0% respectively). Looking at these results, the prediction would be that recovery rates would be lower when ***harder*** mappings were being learnt in the HREO population. It would also provide an opportunity to assess the interaction of timing and recovery by comparing results between the HR population and individuals at a higher risk of early pruning in the HREO population.

A number of distinctions have been made between groups in terms of variability, and parameters that distinguish outcome and recovery rates. However, it is possible that at least some significant outcomes are the product of a noisy learning system. Not all variability can be accounted for with individual parameters. It is highly likely that much of the variability between individuals and groups is caused by higher order inter parameter interactions. That being said, computational models are a useful way to validate theoretical hypotheses, and in

the case of atypicality and recovery, have supported some claims about subgroups and developmental outcomes.

For future work, the first aim is to validate the model against forthcoming empirical outcome data for subgroups now being identified, in order to compare the outcome of non-regressive groups in our model with the early and late subgroups identified in longitudinal clinical research. A priority within clinical research is to understand the implications of autistic subtypes including early onset autism (Landa & Garrett-Mayer, 2006) and the impact of developmental regression (Barger, Campbell & McDonough, 2013). Utilising computational data to investigate pertinent areas of research will be crucial in elucidating the mechanistic differences between subgroups. Identifying more homogenous sub groups of infants with ASD could provide an opportunity for specific, targeted interventions (Charman, 2015). Importantly, computational models provide a basis to explore the potential for behavioural interventions under different assumptions of the cases of deficits. We consider this question in chapter 5.

Chapter 5

5.1 Introduction

As discussed briefly in chapter 1, clinical intervention in autism is an emerging field, and of particular interest is the implementation of early intervention, based on social and communicative divergences in development identified by 12 months. In the current chapter, I build on results from the computational model of ASD analysed in Chapter 4, expanding it to the field of intervention. The model demonstrated that variations in ASD trajectories arose from interactions within the neurocomputational properties of the model, and the putative pathological mechanism, over-pruning. Individuals from a high-risk early onset population were analysed in order to explain the variance in outcome and severity of ASD. Here, I analyse the population after intervention has been initiated.

The chapter examines current findings within intervention research in ASD. I discuss the effectiveness of very early intervention, specifically whether such interventions produce beneficial outcomes. One notable characteristic is the variability in outcome observed following intervention. Of central interest are the factors that predict this variability. I consider higher early IQ levels as a candidate predictor for improvement in overall outcome. I compare the effect of early and late interventions in the model, and the extent to which general ability levels affect intervention results, with the aim of using the model as a complementary approach to clinical intervention data.

Findings from optimal outcome in ASD will also be discussed. One question that stems from this research is whether individuals who "lose" an ASD diagnosis have compensated for early brain changes, or whether changes are caused by 'normalised' patterns of brain activity. I

assess this in the model by analysing two types of intervention. One is aimed at strengthening Regular and Rule verb types in the model at the expense of Irregular verbs, which I will refer to as compensation. The second type is aimed at improving overall performance across the whole training set, which I will refer to as normalisation. I compare the intervention types in order to see whether there is a difference between final intervention scores.

5.2 Early intervention in ASD

As discussed in chapter 1, early interventions in ASD can provide positive outcomes by supporting the development of foundational skills that help with long-term cognitive function. Prospective studies of ASD have reliably identified a number of atypicalities in behavioural and cognitive functioning as early as 12 months that can predict the emergence of ASD. These atypicalities include a decline in IQ over the second year of infancy and atypical sensory and motor behaviours (Bryson et al., 2007). Zwaigenbaum, Bryson and Garon (2013) describe an overall decrease in social-communicative behaviours, such as reduced eye contact and reduced attention towards social scenes (Chawarska et al., 2013). Furthermore, Green, Elsabbagh, Johnson, Charman and Plummer (2014) suggest that differences in parental-infant interactions identified at 9 months are associated with later atypical behaviours such as gaze processing. By 14 months, these atypical behaviours are predictive of an ASD diagnosis. It has been suggested that early perturbations such as these could negatively impact on later cognitive outcomes by affecting existing vulnerabilities that develop during early infancy.

The rationale for the implementation of early intervention stems from a proposal of sensitive periods in development. Brain plasticity is significantly heightened over the first 18 months

of infancy, allowing for accelerated experience-dependent learning (Huttenlocher et al., 2008). This is also the point in development at which initial atypicalities in ASD emerge, and subsequently, atypical developmental trajectories. It is argued that implementing interventions during this sensitive period could alter such trajectories and provide an opportunity to potentially alter long-term developmental outcomes (Webb, Jones, Kelly & Dawson, 2014). This hypothesis is supported by the work of Lewis et al. (2004), who used animal models of ASD to consider to the development of repetitive behaviours. Specifically, a gene-deletion mouse model was implemented to investigate restricted and repetitive behaviours and interests; attributes frequently exhibited in individuals with ASD. They identified a sensitive period of development, known as a "presymptomatic period", at which point intervention could exert a larger influence on emerging behavioural atypicalities. They demonstrated that a phase of early enrichment (rearing the mice in larger and more complex environments) lead to "neuroprotective effects" on a long-term basis. Such effects were measured by showing attenuation in the development of "spontaneous stereotypies" (compared to restricted behaviours in ASD) that developed in mice living in impoverished environments.

Early clinical intervention data supports this hypothesis, and the majority of trials have focused on social and communicative domains, or infant-parental dyadic interactions (Green et al., 2015; Kasari et al., 2014; Wetherby et al., 2014). Interestingly, positive effects have been reported after even brief periods of intervention in early infancy. In a study by Landa and Kalb (2012), 2-year old children diagnosed with ASD received a 6-month evidence-based instruction intervention. Intervention included responsiveness training, which aims to help children to learn and use behaviours that are fundamental to the development of social and cognitive abilities through the use of responsive interactions during daily routines.

Children were enrolled in the nursery-based intervention for 10 hours a week for six months, and additional parental training was provided. Significant gains in IQ and communication were achieved immediately after intervention, and long-term outcomes (37 months post-intervention) showed continued progress. A lack of control group in this study compromises the interpretation of these findings, as the cause of positive outcomes cannot be determined. However, a parent-mediated joint attention based intervention for toddlers with ASD also demonstrated long-term improvements in functional play and joint attention behaviours in comparison to a control group. Furthermore, skills gained from intervention were maintained one year after intervention had ceased (Kasari, Gulsrud, Wong, Kwon & Locke, 2010).

Findings suggest that early intervention may be most effective in terms of cognitive and behavioural outcomes (Webb, Jones, Kelly & Dawson, 2014; Zwaigenbaum et al., 2015). The high-risk infant sibling population is a valuable group that can be used to validate this proposal. Green et al. (2015) reported positive effects of a social communication intervention received between 7 and 10 months, using parental mediated techniques. Interestingly, intervention effects were widespread across cognition and parental-infant dyadic interactions. The authors suggest that this widespread intervention effect supports the notion of a sensitive period of plasticity (see chapter 1 for a review of intervention in ASD).

One striking feature of interventions for ASD is the variability in outcomes across children. While overall effects of intervention are present, the extent of interaction effects from factors such as IQ or environment remains unclear. It is therefore an aim to elucidate the underlying computational parameters in the different intervention populations to generate candidate causal models of how these effects may operate.

5.3 Intervention effects and IQ

A higher IQ is a factor that is associated with positive outcomes in ASD. Howlin et al. (2004) followed up individuals with ASD who had first been assessed at the age of 7. At a follow up (at an average of 29 years of age) individuals with a childhood performance IQ of 70 or above (assessed by measures derived from the Vineland Adaptive Behaviour Scales), demonstrated significantly better outcomes according to standard cognitive measures compared to those with a lower IQ. Similarly, Anderson, Liang and Lord (2013) evaluated children with ASD longitudinally at four time points (2,3,9 and 19 years). A lower IQ (assessed by measures derived by the Mullen Scales of Early Learning) at 2 years was associated with a more severe ASD diagnosis when the same individuals were assessed again at the age of 19. At age 3, IQ was the only significant predictor for inclusion in the low intelligence group for 91% of individuals. Interestingly, both studies concluded that for those in the higher IQ groups, outcome was harder to predict. Within Howlin's groups, IQ was not a consistent predictor in distinguishing between those in "good" and "very good" outcome groups. Anderson et al. (2013), however, reported a prediction rate of remaining in the higher IQ group to be 66% at two years and 82% at three years. The fact that over half of the children in the high IQ population received intervention training, combined with the finding that around 30% of individuals from the lower IQ group improved in terms of adaptive skills substantially over time indicates multiple underlying factors. One might debate the extent to which IQ can be assessed independently of ASD symptoms, and therefore whether IQ in fact represents the severity of ASD prior to intervention. However, these studies used consistent approaches measuring IQ via assessments such as the Mullen Scales of Early Learning and Vineland Adaptive Behaviour Scales, and were able to show separate predictive power of these measures. It will be beneficial to compare with empirical findings, the extent of pre-intervention ability on treatment effects in the model, where the computational basis of IQ is

readily operationalised in terms of the cumulative effect of a parameter set on learning in the absence of pathology.

5.4 Optimal Outcome in ASD

Recent reports of individuals "losing" their ASD diagnosis are challenging the established view that ASD is a lifelong condition. A review of long-term ASD outcomes concluded that between 3% and 25% of individuals with ASD lose their diagnosis over time. However, the phenomenon known as optimal outcome has proved somewhat controversial. It is often assumed that children who lose their diagnosis of ASD were either misdiagnosed and never had autism, or they continued to have autism but exhibited subtler symptomology. Thus far, two studies (Fein et al., 2013; Anderson, Liang & Lord, 2014) have provided evidence that a small subset of individuals have moved out of the autistic spectrum and within typical limits of social, motor and cognitive functioning.

In the study by Fein et al. (2013), initial findings reported a total of 34 individuals demonstrating optimal outcome, defined as no longer meeting the criteria for ASD. Individuals defined as having optimal outcome were not significantly different from typically developing controls in communication or socialisation ADOS assessments. However, three out of the 34 (9%) individuals with optimal outcome exhibited weaknesses on a facial recognition task. The authors argue that 7% of the typical population would be expected to fall at or below 1.5 standard deviations below the mean; therefore, the 9% of individuals with optimal outcome is not below what would be expected by chance in a typical population. Furthermore, when scores from the facial recognition task were compared with a high-functioning autism group, 26% of the high-functioning autism group scored lower than 1.5 standard deviations below the average, demonstrating a significantly higher impairment level

as a group. The authors argue that this result substantiates the possibility of optimal outcome in ASD. In the second study, Anderson, Liang and Lord (2014) reported results from 85 children with autism that had been followed longitudinally between 2 and 19 years of age. Of the 32 individuals that did not demonstrate intellectual disability, eight children (9%) no longer met the diagnostic criteria for ASD. The authors called this a "very positive outcome".

A question arising from optimal outcome research is whether subtle deficits are present in areas of social and communicative interactions or cognition, or whether brain structure and function have (to some extent) normalised. Dawson et al. (2012) have shown evidence of normalised brain activity patterns after intervention. Children receiving behavioural interventions in the Early Start Denver Model demonstrated multiple cognitive and social behavioural improvements, and normalised neural patterns when viewing faces. Conversely, Fein at al. (2013) suggest that optimal outcome individuals generally exhibited higher IQ levels. They consider the idea that this higher IQ could be a demonstration of compensatory mechanisms (substituting weak social processing for explicit processing for example) rather than normalisation. Fein and colleagues also produced brain-imaging data to support the idea that brain function is not normalised in individuals with optimal outcome. In their unpublished study, they reported that individuals in the optimal outcome group produced patterns of neural activity in a language task that appeared to resemble those of individuals with ASD rather than patterns seen in typically developing individuals. This would suggest that compensatory neural mechanisms might be underlying what is seen as typical cognition in individuals with optimal outcome.

Thus far, attempts at understanding and utilising intervention in ASD have shown limited

success. Studies have hypothesised that early intervention may be beneficial in order to initiate interventions at a time when the brain is in a state of optimal plasticity. Furthermore, a small number of individuals have demonstrated remission later in development. What is still unclear, however, is the cause of variability in response to intervention. It must be questioned why some individuals benefit from intervention sometimes to the point of losing an ASD diagnosis, whilst others exhibit little or no change. Currently, the significance of IQ, environment, timing of intervention and effects of intervention on later outcomes remain unclear.

5.5 Modelling intervention effects

The implementation of clinical intervention trials is expensive, time consuming and must be rigorously evaluated before it can be utilised in infants. Computational models offer a unique opportunity to clarify terminological distinction through the implementation of different intervention types. Unlike clinical intervention trials, a single population can be utilised in multiple models, allowing for the comparison of intervention effects across the same individual networks. Furthermore, variability within individual networks is accounted for by the computational parameters included in each iteration of the model. Therefore, it provides the potential to generate possible candidate mechanisms to account for the variability in response to intervention and validate putative theoretical claims.

In this chapter, I identify two possible intervention sets to apply to disordered networks. The aim of first intervention set was to 'normalise' behaviour. The aim of the second intervention set was to compensate for an atypical learning system, and optimise performance levels. I compared intervention effects from normalisation and compensatory training sets in the model in order to elucidate effects of timing of intervention, and type of intervention.

Specifically, I assessed the importance of the timing on intervention, specifically whether larger intervention effects were exhibited in networks with an earlier intervention onset, and the duration of positive intervention effects across development. Finally, I identified neurocomputational parameters predicting individual response to intervention.

5.6 Method

5.6.1 Design

As discussed in Chapter 4, this study took advantage of a pre-existing computational model by Thomas, Knowland, and Karmiloff-Smith (2011). The model hypothesised that over-pruning of neural networks and subsequent connectivity could be an underlying mechanistic cause of ASD. Results from Chapter 4 suggest that variations in ASD trajectories can arise from the same pathological (over-pruning) mechanism due to, on the one hand, differences in severity, and on the other hand, interaction of the pathology with population-wide individual differences in other neurocomputational parameters. The model therefore accounted for multiple types of ASD: early onset, late onset, regressive subtypes and the broader autism phenotype (BAP) through the implementation of one pathological mechanism.

The original model considered populations comprised of 1000 individual networks trained to learn English past tense forms of varying degrees of difficulty and irregularity. Fourteen neurocomputational parameters and a parameter used to manipulate the learning environment (known as the family quotient) were used to create variability within the populations (see Chapter 4 for a full description of all neurocomputational parameters). These parameters varied across networks and each individual was trained for a total of 1000 epochs (where one epoch was an exposure to all patterns in an individual training set). The model considered three populations: The first was low-risk population, where individual networks were

provided with a reasonable learning environment (with family quotient levels between 0.6 and 1) and a low pruning threshold (with a maximum level of 1.5 and an average value of 0.1). The second population was at a higher risk of being subjected to aggressive pruning. The pruning threshold in this group ranged between 0.75 and 4.0 and significant effects of pruning were more likely to occur than in the low-risk group. The third population was the low-risk deprived population. All parameters were varied in the same range as the low-risk population apart from the family quotient, which varied between 0 and 1 (compared to 0.6 and 1 in the low risk population). This increased the risk of a network being exposed to an impoverished learning environment.

As we saw in Chapter 4, for this thesis, two additional simulated populations were created. These were high-risk early onset (HREO), and high-risk early onset control (HREOC). All parameters in the HREO population were sampled from the same distributions as the original high-risk population apart from the pruning onset parameter, which was set to increase the likelihood of early onset pruning. The HREOC differed from the HREO only in that a pruning threshold was set to a lower value (where the average pruning threshold was 0.1).

The HREO group exhibited a range of atypical trajectories, which provided an appropriate foundation to explore the effect of intervention. Here, intervention was simulated according to four variables: (1) the selected intervention set, (2) the epoch of training at which intervention was applied, (3) the duration in epochs that intervention persisted, and (4) the dosage of intervention in terms of the proportion of intervention items to normal training items. The analyses focus on three aspects: the type of intervention (normalisation versus compensation); the effect of timing in intervention (compensatory early versus compensatory late); and the factors that predicted variable response to intervention.

5.6.2 Construction of intervention sets

Two types of intervention set were created. Each comprised a number of artificial verbs consistent with the training domain of the base past-tense model utilised by Thomas, Knowland and Karmiloff-Smith (2011). I assumed that the behavioural intervention was much smaller in scale than continued everyday experience, and so limited the intervention set to less than 10% the size of the original training set (42 input-output mappings versus 508 in the original set). The method followed one of the broad tenets of an intervention drawn from speech and language therapy called grammar facilitation, a widely investigated intervention method used to address grammar deficits in school age children. In grammar facilitation, the aim is to make target forms more frequent. This is hypothesised to help the child identify grammatical rules and give them practice at producing forms they tend to omit (Ebbels, 2014). In line with this approach, the intervention added information to the training set model for a fixed period, to increase the salience of certain regularities in the problem domain.

The normalisation set was designed to improve performance on the full training set. Three past tense models with 'typical' parameter values (most frequent in the population) were trained using different random seeds. The networks employed a 3-layer architecture with a single layer of hidden units. At the completion of (successful) training after 1000 epochs, the hidden unit activations were visualised according to the first and second principal components across the full training set. This gave three 2-dimensional plots of well-structured hidden unit representations. The rationale was to select a subset of items from the training set that spanned this space, to encourage networks to adopt a well-formed representational solution. For each of the plots from the three successfully trained networks, seven regular verbs, three no-change irregulars, three vowel-change irregulars, and 1

arbitrary irregular were selected spanning the similarity space, summing to 42 existing training patterns overall.

The compensation set was designed to improve generalisation of the past tense rule, potentially at the expense of performance on irregular verbs that violate this rule. Construction of the compensation set utilised a set of 500 novel verbs generated by Yang and Thomas (2015). These were optimised in terms of their phonological features to generate accurate regular past tense formation, using the same coding scheme and architecture. From this set of novel verbs (Yang and Thomas, 2015; Voice Feature Satisfied data set), the same procedure was followed as for the above normalisation set. The 500 verbs were applied to the three successfully trained networks, and their positions in hidden unit space were plotted according to the same first principal component dimensions identified for the training items (i.e., using the eigenvectors of the PCA carried out using the training set). For each of the three plots, 14 novel verbs were identified that spanned the similarity space. This summed to a compensation set of 42 novel training items overall.

5.6.3 Procedure for simulating intervention

A behavioural intervention to remediate the developmental impairment was simulated in the following way. It was assumed that the impairment was diagnosed at some point relatively early in development. Two points of intervention were contrasted: 30 epochs and 100 epochs to evaluate the possible implication of differences in timing of intervention. For back propagation networks, it has been argued that plasticity reduces across training (see Thomas & Johnson, 2006, for discussion). One mechanism producing this reduction is the loss of network connections, making it a pertinent dimension to explore given the Over-Pruning hypothesis.

At the point of intervention, items were added to the original training set. The intervention set added to, rather than replaced the original training set under the assumption that in a clinical setting, interventions take place against the child's continued experience of his or her normal learning environment. In addition, for an artificial neural network, replacement would run the risk of catastrophic interference. Intervention continued for a limited duration, after which the intervention ceased and training reverted to the original set. Following piloting, the duration was set at 40 epochs. During normal training, the token frequency of training items was implemented according to a constant that modified the size of the weight change during learning (see Plaut et al., 1996). High frequency items were given the constant of 0.3, while low frequency items were given the constant of 0.1, so that high frequency items had a greater impact on the alteration of network weight matrices. This was an implementation decision to speed up learning (see Plaut, McClelland, Seidenberg & Patterson, (1996). Items in the intervention set were given a constant of 0.4, to reflect their higher salience for the child. This defined the dosage of the intervention.

5.6.4 Dependent variables and predictors of response to intervention

The design contrasted two intervention sets (normalisation versus compensation) and two timing points for intervention (30 epochs versus 100 epochs). Performance was assessed using five types of verb learning in the past tense domain that increased in difficulty level. As in Chapter 4, the five performance levels used in the training set were known as Regular, Rule (generalisation), and three irregular metrics, which increased in difficulty level, known as Hard, Harder and Hardest learnt. See Chapter 4 for a full explanation of the performance metrics and training set used in the model.

Normalisation-based intervention considered the complete set of performance metrics, whereas in compensatory-based intervention, the focus was on regular verb performance and the generalisation of a common rule since irregular verb performance would by definition be expected to decline. Intervention itself was measured by a change in accuracy level in verb acquisition in comparison with the same, untreated version of an individual network. Using simulation data to examine intervention effects provided a unique advantage to examine the outcome of both treated and untreated systems, with all other parameters held equal. Intervention could be assessed at one of two points: either at the end of the fixed intervention duration (70 epochs for early, or 140 epochs for late intervention groups), or at the end of training altogether (1000 epochs).

In Chapter 4, Regular mappings were the focus of all analyses. In the current chapter, Regular, Rule and Irregular mappings were analysed. One of the aims of Chapter 4 was to compare three computational populations to assess the relationship between timing, onset and severity of pruning on the development of both regressive and non-regressive networks. A second aim was to identify non-regressive subgroups based on development. Due to the size of the datasets and the research questions, it seemed most appropriate to focus on the predominant mappings in the datasets, which allowed for a broader behaviour of networks. In the current chapter, both the HREO population as a whole and the identified subgroups are compared across two types of intervention. The first intervention used prototypical exemplars from the full training set with the aim of normalising behaviour across the entire training set (normalisation). The second intervention aimed at compensating the learning system to optimise the acquisition of Regular and Rule (generalisation) mappings at the expense of the irregular verbs in the training set. Unlike Chapter 4, the research questions here utilise all verb types; it is therefore essential to analyse the full training set here.

Given that intervention was applied on a population with simulated individual differences, a number of networks are likely to exhibit poorer performance levels due to the limits of their own family training set. This is particularly likely with irregular or exception verbs (that is, Hard, Harder and Hardest), as prior knowledge of other known words would not benefit a network when it was required to 'guess' novel items. For the normalisation intervention training set, one might predict an overall improvement in learning irregular mappings in some cases, simply because the network had not yet encountered this item. It is therefore essential to interpret such results with caution

Response to intervention was assessed across the both the HREO population as a whole, and across groups that were identified in Chapter 4. Inspecting the difference in the slope of each individual developmental trajectory identified four non-regressive subgroups. These were known as Messy, Lower, Slower and Typical. A regressive subgroup of the HREO population was also identified. Predictors of response to intervention were considered within each group. Predictors included the computational parameters of each individual network and each network's family quotient, which determined the richness of the environment in the training set.

As previously mentioned, there have been suggestions from empirical intervention literature that IQ might be predictive of some of the variability in children's responses to intervention in ASD (Howlin, Magiati & Charman, 2009; Fein et al., 2013). How might we capture the 'IQ' of a network? Here, IQ is referring to the capabilities of a network as a whole. This is the net effort of all computational parameters on learning ability. One measure that would index this effect is the performance level prior to the beginning of intervention, averaged

across all verb types.

Alternatively, one could attempt to analytically derive the combined effect of all its neurocomputational parameters on its ability to learn the training set. It would be advantageous to consider this as separate from pathology, thereby indexing the 'reserve' of the network to recover despite pathology. Here, we separate IQ from pathology in the same way that in prospective studies it would be possible to measure the IQ of a high-risk infant before the symptoms of ASD have developed, or estimate it based on family background (e.g. unaffected sibling IQ or averaged parental IQ). In the model, we can measure a network's intrinsic ability level in the absence of pruning.

A complete measure of learning ability was too complex to derive formally, due to the possible higher order interactions between the multiple parameters varying in each network. Behaviourally, learning ability could only be assessed once some period of development had taken place, but such a measure would then be contaminated by variation in the particular environment to which the network had been exposed. This problem was solved in the following way. For the HREOC population (with pruning threshold parameter set to 0.1 for all networks), population performance was taken on one of the more discriminating performance measures, vowel-change irregular verbs after 100 epochs of training. A multiple linear regression analysis was carried out to predict network performance from the neurocomputational parameters and family quotient value. The statistical model explained around 40% of the population variance and yielded a linear equation of weighted parameter values that best predicted each individual's performance. The linear equation was then combined with each network's parameter values – excluding their family quotient – to generate a score that represented the network's intrinsic learning ability *independent* of the

environment and *in the absence* of pathology. This could then be used as a predictor of response to intervention.

5.7 Results

First, I compared treated and untreated versions of each network in order to assess intervention effects on the whole population. Treatment effects were generated for the full population by comparing each treated network with its untreated version. Positive treatment effects arose where performance was higher in the treated version of the network; negative treatment effects arose where performance was lower in the treated network.

Effects of timing, intervention type and measurement point were contrasted for each of the five dependent variables using three-way ANOVAs. I consider which factors predict individual response to intervention, and elucidate the causes of variability in response to treatment effects using multiple regression models. Due to the vast amount of possible regression analyses that could have been performed, I analysed only those that demonstrated significant intervention effects. Initially, all computational parameters were included as predictor variables. Two further predictors were computed to indicate the IQ or ability level of individual networks.

The results are split into two sections. The first half of the analyses compare mean treatment effects per intervention, population, subgroup and measurement type. The second half explores the predictors of individual variation in response to intervention.

5.7.1 Effect of intervention

First, paired t-tests compared untreated (UN) against treated (TR) overall scores across the three intervention populations. Scores were compared across all difficulty measures before intervention (30 epochs for early intervention, 100 epochs for late intervention), at the end of

intervention (70 of 100 epochs respectively), and the end of training (1000 epochs). Tables 5.1-5.5 provide a full overview of the results, indicating both the size of intervention effects and their modulation by various manipulations (group, intervention type, measure, measurement point, and timing). Each table represents the results for an individual measurement type.

Table 5.1. Intervention effects in the Regular measurement condition.

* Independent t-test treated vs. untreated p<.05

+Independent t-test treated vs. untreated p<.01

Group		Accuracy for groups intervened upon at an early (30 epochs) or late (100 epochs) time point								
		Compensatory Early			Compensation Late			Normalisation Early		
		30	70	1000	100	140	1000	30	70	1000
Overall	UN	0.603	0.643	**+0.671**	0.607	0.631	0.671	0.603	0.643	0.671
	TR	0.603	0.650	**0.678**	0.607	0.596	0.652	0.603	0.617	0.685
Messy	UN	0.637	0.733	0.797	0.744	0.767	0.797	0.637	0.733	0.797
	TR	0.637	0.723	0.797	0.744	0.728	0.793	0.637	0.640	0.804
Low	UN	0.424	0.495	0.654	0.528	0.556	0.654	0.424	0.495	0.654
	TR	0.424	0.504	0.653	0.527	0.526	0.635	0.424	0.456	0.664
Slow	UN	0.457	0.632	0.834	0.687	0.724	0.834	0.457	0.632	0.834
	TR	0.457	0.640	0.833	0.687	0.706	0.832	0.457	0.600	0.842
Typical	UN	0.767	0.869	0.937	0.891	0.907	0.937	0.767	0.869	**+0.937**
	TR	0.767	0.869	0.938	0.891	0.897	0.936	0.767	0.855	**0.940**
Regressive	UN	0.577	**+0.573**	**+0.555**	0.499	0.523	0.555	0.577	0.573	**+0.555**
	TR	0.577	**0.583**	**0.565**	0.499	0.478	0.527	0.577	0.549	**0.574**

Table 5.2. Intervention effects in the Rule measurement condition.

* Independent t-test treated vs. untreated / control vs. untreated p<.05

+ Independent t-test treated vs. untreated / control vs. untreated p<.01

Group		Accuracy for groups intervened upon at an early (30 epochs) or late (100 epochs) time point								
		Compensatory Early			Compensation late			Normalisation Early		
		30	70	1000	100	140	1000	30	70	1000
Overall	UN	0.487	**0.52**	**0.547**	0.499	0.522	0.547	0.487	0.52	0.547
	TR	0.487	**0.544**	**0.561**	0.499	0.185	0.54	0.487	0.486	0.555

Messy	**UN**	0.523	0.605	0.64	0.608	0.631	0.64	0.523	0.605	0.64
	TR	0.523	0.623	0.656	0.608	0.623	0.656	0.523	0.514	0.641
Low	**UN**	0.361	0.429	0.549	0.454	0.477	0.549	0.361	0.429	0.549
	TR	0.361	0.451	0.558	0.454	0.466	0.547	0.361	0.387	0.549
Slow	**UN**	0.381	**0.514**	**0.623**	0.552	0.574	0.623	0.381	0.514	0.623
	TR	0.38	**0.538**	**0.641**	0.553	0.581	0.64	0.381	0.475	0.626
Typical	**UN**	0.605	**0.664**	**0.679**	0.671	**0.677**	**0.679**	0.605	0.664	**0.679**
	TR	0.605	**0.69**	**0.699**	0.671	**0.695**	**0.695**	0.605	0.637	**0.683**
Regressive	**UN**	0.469	**0.474**	**0.487**	0.432	0.46	0.487	0.468	0.474	**0.487**
	TR	0.469	**0.497**	**0.5**	0.431	0.43	0.468	0.468	0.441	**0.499**

Table 5.3. Intervention effects in the Irregular Hard measurement condition.

***** Independent t-test treated vs. untreated / control vs. untreated p<.05

+ Independent t-test treated vs. untreated / control vs. untreated p<.01

Group		**Accuracy for groups intervened upon at an early (30 epochs) or late (100 epochs) time point**								
		Compensatory			**Compensation late**			**Normalisation early**		
		Early								
		30	**70**	**1000**	**100**	**140**	**1000**	**30**	**70**	**1000**
Overall	**UN**	0.305	**0.341**	0.366	0.311	0.315	**0.366**	0.305	0.341	0.366
	TR	0.305	**0.351**	0.368	0.311	0.314	**0.387**	0.305	0.603	0.408
Messy	**UN**	0.3	0.353	0.423	0.357	0.348	0.423	0.3	**0.353**	0.423
	TR	0.3	0.3	0.208	0.357	0.357	0.394	0.3	**0.631**	0.434
Low	**UN**	0.184	0.213	0.282	0.212	0.235	0.282	0.184	0.213	0.282
	TR	0.184	0.2	0.265	0.212	0.197	0.229	0.184	0.449	0.312
Slow	**UN**	0.189	0.273	0.581	0.316	0.386	0.581	0.189	**0.273**	**0.581**
	TR	0.189	0.268	0.576	0.316	0.335	0.552	0.189	**0.614**	**0.622**
Typical	**UN**	0.41	0.563	0.816	0.636	0.684	0.815	0.41	**0.563**	**0.815**
	TR	0.41	0.576	0.815	0.636	0.661	0.805	0.41	**0.824**	**0.856**
Regressive	**UN**	0.294	0.283	0.188	0.206	0.186	0.188	0.294	0.283	0.188
	TR	0.294	0.299	0.169	0.206	0.201	0.233	0.294	0.534	0.233

Table 5.4. Intervention effects in the Irregular Harder measurement condition.

***** Independent t-test treated vs. untreated / control vs. untreated p<.05

+ Independent t-test reated vs. untreated / control vs. untreated p<.01

Group		Accuracy for groups intervened upon at an early (30 epochs) or late (100 epochs) time point								
		Compensatory Early			Compensation Late			Normalisation Early		
		30	70	1000	100	140	1000	30	70	1000
Overall	UN	0.151	0.225	0.285	0.19	0.205	0.285	0.151	**0.225**	0.285
	TR	0.151	0.218	0.286	0.19	0.185	0.275	0.151	**0.292**	0.308
Messy	UN	0.095	0.147	0.231	0.158	0.174	0.231	0.095	**0.179**	0.231
	TR	0.094	0.117	0.208	0.158	0.131	0.195	0.095	**0.239**	0.263
Low	UN	0.025	0.056	0.115	0.067	0.076	0.115	0.025	**0.072**	0.115
	TR	0.024	0.054	0.119	0.066	0.064	0.103	0.025	**0.13**	0.134
Slow	UN	0.047	0.106	0.484	0.146	0.201	0.484	0.047	0.153	**0.484**
	TR	0.046	0.097	0.482	0.146	0.179	0.463	0.047	0.258	**0.52**
Typical	UN	0.24	0.457	0.795	0.547	0.624	0.795	0.24	**0.582**	**0.795**
	TR	0.24	0.441	0.796	0.547	0.58	0.787	0.24	**0.667**	**0.811**
Regressive	UN	0.146	0.178	0.1	0.086	0.076	0.1	0.146	**0.216**	**0.1**
	TR	0.145	0.174	0.103	0.086	0.063	0.092	0.146	**0.27**	**0.122**

Table 5.5. Intervention effects in the Irregular Hardest measurement condition.

* Independent t-test treated vs. untreated / control vs. untreated p<.05
+ Independent t-test treated vs. untreated / control vs. untreated p<.01

Group		Accuracy for groups intervened upon at an early (30 epochs) or late (100 epochs) time point								
		Compensatory Early			Compensation Late			Normalisation Early		
		30	70	1000	100	140	1000	30	70	1000
Overall	UN	0.218	0.342	0.335	0.276	0.283	0.335	0.29	0.342	0.402
	TR	0.218	0.313	0.336	0.276	0.227	0.32	0.29	0.403	0.471
Messy	UN	0.248	0.342	0.526	0.377	0.387	0.526	0.248	**0.309**	0.526
	TR	0.248	0.245	0.49	0.377	0.223	0.484	0.248	**0.396**	0.497
Low	UN	0.044	0.1	0.212	0.127	0.127	0.212	0.044	**0.084**	0.212
	TR	0.044	0.056	0.188	0.127	0.059	0.176	0.044	**0.14**	0.235
Slow	UN	0.101	0.217	0.589	0.298	0.378	0.589	0.101	**0.169**	**0.589**
	TR	0.101	0.174	0.6	0.298	0.254	0.574	0.101	**0.279**	**0.602**
Typical	UN	0.458	0.686	0.794	0.74	0.763	0.794	0.458	**0.561**	0.794
	TR	0.458	0.656	0.793	0.74	0.709	0.79	0.458	**0.624**	0.796
Regressive	UN	0.276	0.26	0.145	0.122	0.115	0.145	0.276	0.222	0.145
	TR	0.276	0.236	0.149	0.122	0.072	0.128	0.276	0.276	0.168

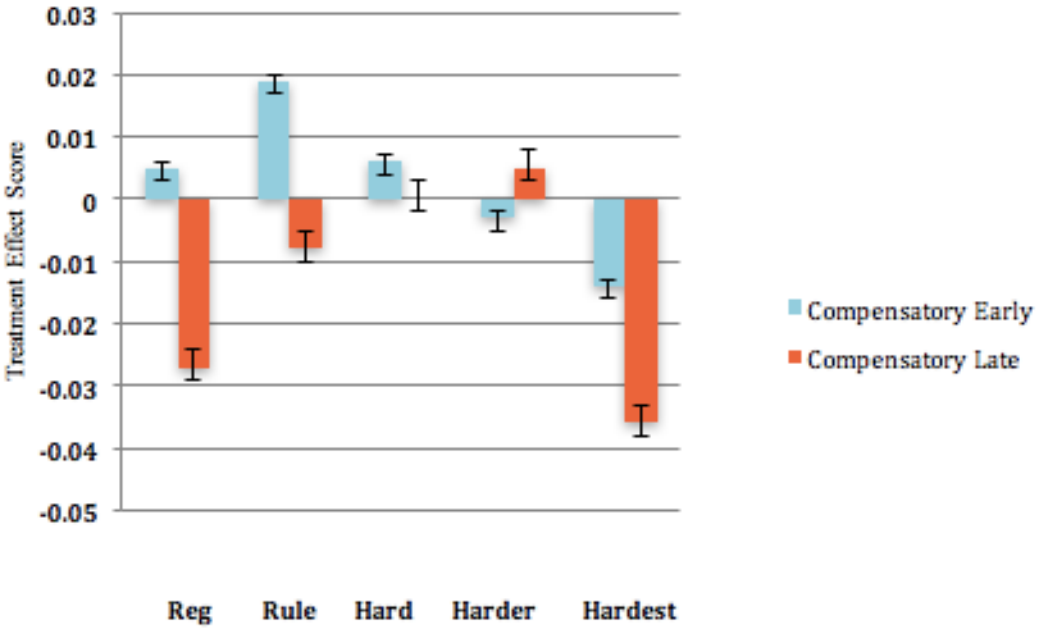
Both compensatory and normalisation groups demonstrated positive intervention effects across Regular, Rule and Hard measures. Stronger effects were demonstrated when

intervention was implemented earlier (30 epochs) rather than late (100 epochs). In terms of subgroups identified in chapter 4, Slower and Typical groups showed larger treatment effects. Overall however, all intervention effects were small. The difference between the HREO population and its control condition provided a benchmark of the size of behavioural deficits produced by pruning.

5.7.2 Effect of timing

It was of interest to explore whether treatment effect size depended on the time at which intervention was delivered. A three-way ANOVA was run, whereby timing (compensatory early or compensatory late) measurement type (Reg, Rule, Hard, Harder and Hardest) and measurement point (end of intervention or end of training) were included as repeated measures. A significant main effect of timing was found $F(1, 999) = 110.703$, $p < .05$. As predicted, mean scores for late implementation of intervention were significantly lower (M = -.013) than those in the early intervention populations (M = .003). The only significant interaction showed that the effects of timing were modulated by measurement type. Figure 5.1 shows the mean treatment effect scores. Early intervention was beneficial in the Regular and Rule measurements, whereas compensatory conditions show limited positive intervention effects. The positive mean effect in the Harder condition is likely to be masking a large amount of variability within the individuals in the late population.

Figure 5.1 Comparing the mean treatment effect scores for timing of intervention across the five measurement points. Error bars show the standard error of the mean.

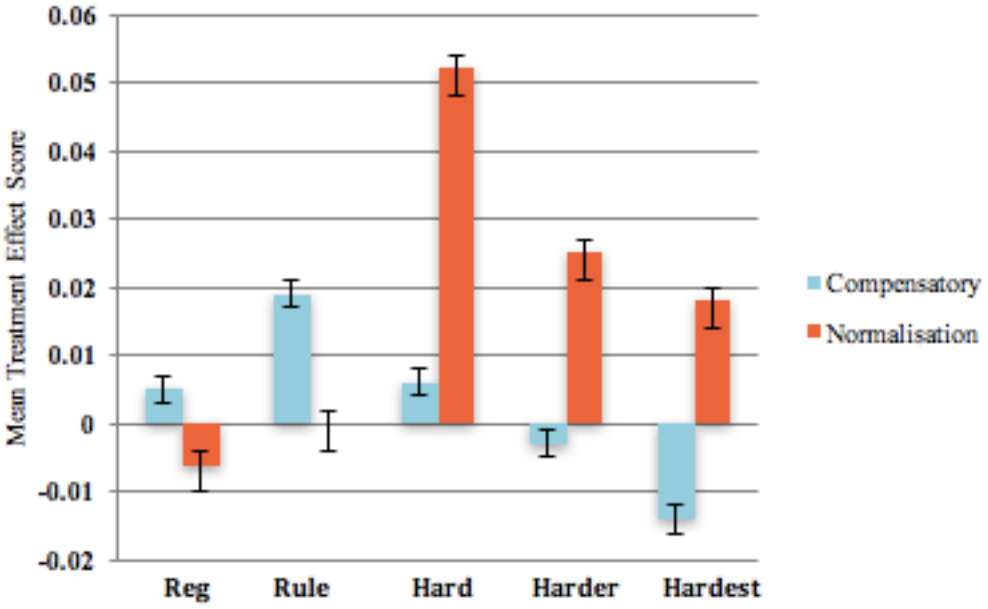


5.7.3 Intervention type

A three-way repeated measures ANOVA was run to identify the extent to which intervention effect depended on intervention type. Therefore, the three repeated measures were the same as for the effect of timing; however, the timing measurement was replaced by the intervention type measure (Compensatory Early or normalisation Early). A main effect of intervention type was demonstrated $F(1, 999) = 148.569$, $p < .05$. Mean intervention effects were positive in both normalisation (M = .003) and compensatory groups (M = .018), and overall, the normalisation population displayed marginally higher treatment effects. The only significant interaction effect was shown between intervention type and measure type $F(2.34, 3996) = 127.26$, $p < .05$. Figure 5.2 shows the mean treatment effects for the normalisation and Compensatory populations across the five measurement points. Compensatory intervention was more successful in the Regular, Rule and Hard conditions, whereas for the normalisation intervention set, treatment effects were positive only in the irregular (Hard, Harder and

Hardest) conditions.

Figure 5.2. Comparing mean treatment effect scores for intervention population types across the 5 measurement points. Error bars show the standard error of the mean.



5.7.4 Effect of group

Three-way mixed ANOVAs were also used to ascertain whether the type of subgroup affected treatment effects of intervention. A total of three ANOVAs were computed, representing one per population type (compensatory early, compensatory late, normalisation early). In each case, measurement point, with two levels (end of intervention vs. end of training) and measure type, with five levels, (Regular, Rule Hard, Harder and Hardest) were included as repeated measures, and subgroup was included as a between subjects factor. A main effect of subgroup was present in the compensatory Early population $F(1, 991) =$

10.242, $p < .05$. Overall treatment effects were positive only in the Typical (M = .001) and Regressive (M = .006) subgroups. No significant interaction effects were identified.

In the Compensatory Late intervention group, a main effect of subgroup was also present, $F(4, 993) = 7.229$, $p < .05$. Messy networks demonstrated the lowest overall intervention effects, followed by Lower, Slower and Regressive. The Typical subgroup demonstrated the highest overall score. There was a significant interaction between the measure type and subgroup $F(16, 3972) = 14.55$, $p < .05$. Estimated marginal means showed that the only interaction with positive effects was the Rule measure in the Messy (M= .001) Slower (M = .004) and Typical (M = 0.03) conditions. Conversely, Harder and Hardest irregular exception measures negatively affected treatment effects across all groups. The effect of measurement point on subgroup was non significant $F(4, 993) = 2.30$, $p = .057$.

The normalisation population also demonstrated a main effect of subgroup group $F(4, 987) = 4.853$, $p <.05$. The Messy and Lower subgroups showed the lowest mean intervention effects. However, unlike the compensation early populations, all group mean effects were positive values, and therefore demonstrated positive intervention effects overall. The group with the highest treatment effect scores was the Slower subgroup. A significant interaction between measurement type and subgroup was demonstrated $F(16, 2016) = 2.673$, $p < .05$, and a significant interaction between subgroup and measurement point was also significant $F(4, 987) = 12.81$, $p < .05$. All subgroups had higher mean effects at the end of intervention compared to the end of training. However, at the end of training, all subgroups had maintained positive intervention effects.

5.7.5 Individual variability

Frequency distributions were used to plot significant intervention effects. Two frequency distributions are shown here, to illustrate the fact that small mean effects masked a lot of individual variability. The frequencies below were taken from the Typical and Slower subgroups in the compensatory early populations. Regression analyses were conducted to elucidate the cause of variability among individual networks.

Figure 5.3 Frequency distribution for the Slower subgroup at 70 epochs using the Rule measurement in the compensatory early population. The Y-axis shows change in the proportion correct in application of the regular (+ed) suffix to novel verbs.

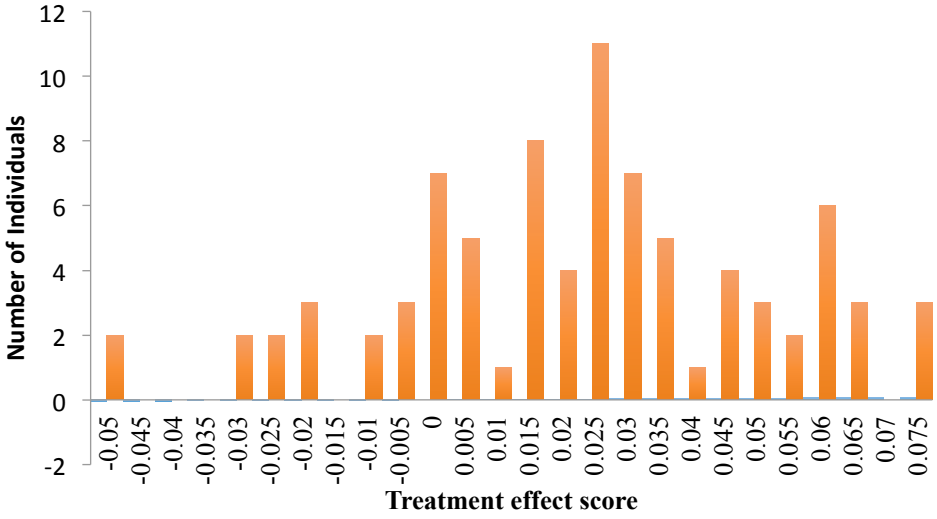
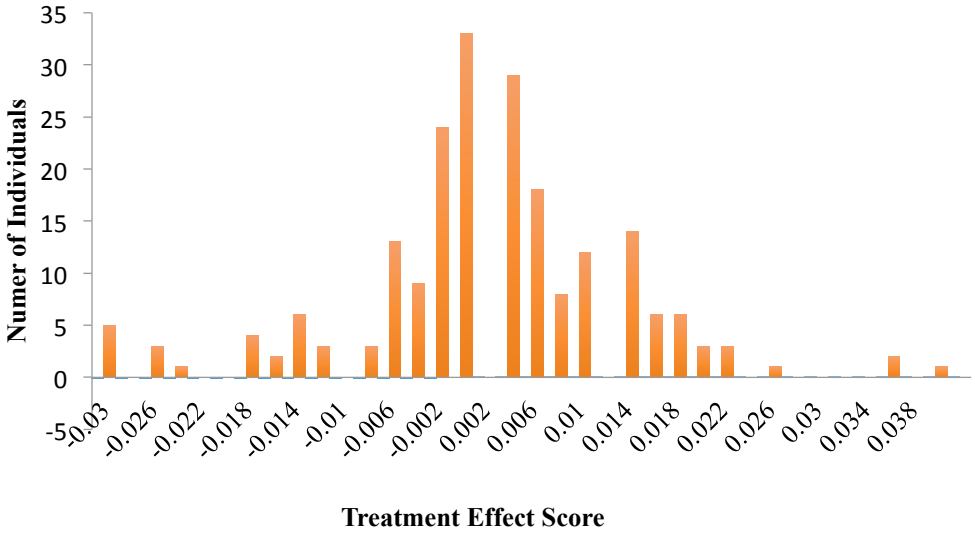


Figure 5.4 Frequency distribution for the Typical subgroup at 1000 epochs using the Regular measurement in the compensatory early population. The Y-axis shows change in the proportion correct in application of the regular (+ed) suffix to novel verbs.



5.7.6 Individual differences in response to intervention

Multiple linear regression models were run on all significant population groups. It was of interest to identify the neurocomputational parameters predicting individual responses to intervention, and understand whether variability in response to intervention was caused by the severity of the pathology, or interplay of neurocomputational parameters fluctuating across intervention populations. The results have been split into the three intervention population types and all regression analyses used treatment effect data. This score represents the total untreated network scores, minus the total treated network scores after intervention. The regressions represent statistically significant scores. Groups with non-significant treatment effects were not included for further analysis, and any non-significant regression models were not reported (see appendix 5, showing the correlations between predictors). In the following analyses, I include typically developing networks. I wanted to establish which parameters predict response to intervention and was initially neutral as to whether the presence of pathology is modulating their effect, as one might be neutral as to whether an intervention targeted to improve, say social skills in a child with ASD might enhance skills in

a typically developing child.

5.7.7 Predicting individual variation in compensatory early population

End of intervention Regular performance intervention effects

A total of three multiple regression models reliably predicted overall intervention effects from the neurocomputational and pre-intervention ability parameters. Firstly, intervention effects at the end of intervention, for the Regular measurement group significantly predicted intervention effects $F(15,984) = 12.162$, $p < .01$, adjusted $R^2 = .15$. Three significant predictors were identified, and the relative contribution of each predictor to the total variance; Ability Pre-Intervention, $\beta = .13$, $t(986) = 1.25$, $p = .03$ (2%), Temperature $\beta = .02$, $t(986) = 7.28$, $p = .012$ (10%), and Architecture $\beta = -.27$, $t(986) = -2.21$, $p < .01$ (2%).

End of training Regular performance

Intervention effects were also predicted at the end of training using the Regular performance measure $F(13,986) = 2.034$, $p < .01$, adjusted $R^2 = .13$. The significant predictor in the model was Pruning Threshold $\beta = .005$, $t(986) = 2.17$, $p < .01$ (5%).

End of training Rule performance

A third multiple regression model significantly predicted intervention effects at the end of training that implemented the Rule measure $F(111,986) = 2.441$, $p < .05$, adjusted $R^2 = .02$. The significant predictors of the model were Pruning Probability $\beta = -.017$, $t(986) = -.85$, $p = .036$ (0.9%) and Ability Pre-Intervention $\beta = .083$, $t(986) = -.54$, $p < .01$ (0.6%). End of training and regression models using the Irregular Hard performance measure were non-significant.

Multiple regression models were also run for across the five subgroups in order to compare the underlying parameters response for the variation. Significant intervention effects were demonstrated in the Typical group, using the Rule measurement at the end of training $F(16,546) = 1.716$, $p = < .05$ adjusted $R^2 = .12$. Hidden Units $\beta = .009$, $t(986) = 2.89$, $p < .01$ (3%), and Architecture $\beta = .09$, $t(986) = 1.56$, $p < .01$ (2%) were significant predictors.

5.7.8 Predicting variation in the compensatory late population

In terms of overall analysis of the compensatory late intervention group, one regression model significantly predicted intervention effects. The model compared the Hard measurement at the end of training $F(13,986) = 10.320$, $p < .01$, adjusted $R^2 = .10$. Ability Pre-Training $\beta = .010$, $t(986) = 1.12$, $p < .01$ (2%), General Ability $\beta = .001$, $t(986) = .044$, $p = 0.03$ (0.6%), Hidden Units $\beta = .001$, $t(986) = 0.002$, $p < .01$ (0.5%) and Pruning Threshold $\beta = -0.03$, $t(986) = 2.15$, $p < .01$ (2%) were all significant predictors.

All subgroup analyses were non-significant in the compensatory late group. Compensatory early populations look comparatively different, with pre-intervention ability also affecting intervention. The direction is as expected; a higher pre-intervention score predicted larger treatment effects.

5.7.9 Predicting individual variation in normalisation early population

In the normalisation early intervention population, there were two significant models that predicted intervention effects. End of training implementing the Rule measurement significantly predicted such effects $F(13,986) = 1.99$, $p < .05$, adjusted $R^2 = .04$. One significant predictor, Pruning Threshold was identified $\beta = -.004$, $t(986) = -1.35$, $p = .046$ (0.8%).

The end of intervention under the Hard measurement also predicted positive intervention effects $F(13,986) = 4.515$, $p < .01$, adjusted $R^2 = .143$. Pruning Threshold was again the only significant predictor $\beta = -.03$, $t(986) = -3.151$, $p < .01$ (4%). These results suggest that the pruning pathology is influencing the intervention effect on normalisation conditions in comparison to population wide variation in other parameters.

Looking at subgroups in the normalisation early intervention population, a regression model significantly predicted intervention effects in Slower subgroup, using the Regular measurement, at the end of intervention $F(15,74) = 5.681$, $p < .01$, adjusted $R^2 = .24$. Hidden Units $\beta = .30$, $t(986) = -8.255$, $p < .01$ (9%), Learning Rate $\beta = .034$, $t(986) = 2.34$, $p < .01$ (3%), Pruning Threshold $\beta = -0.12$, $t(986) = -2.53$, $p < .01$ (3%) and General Ability $\beta = .118$, $t(986) = 2.04$, $p < .01$ (3%) were significant predictors.

The model also predicted intervention effects in the normalisation early population in the Typical group, using the Regular measurement at the end of training $F(15,194) = 1.818$, $p < .05$ adjusted $R^2 = .05$. Pruning Threshold $\beta = -.283$, $t(986) = 3.22$, $p < .01$ (4%) was the only significant predictor.

As expected from results in Chapter 4, the subgroups showing significant intervention effects are consistently the Typical and Slower groups. The highest mean scores in each population were from networks in the Typical subgroup. This is concurrent with findings in Chapter 4, whereby Typical and Slower groups were able to make a greater recovery from pruning in comparison to the Messy and Lower subgroups. An interesting finding was that the majority of the total amount of variance explained was not predicted by the included variables, even

though, for example, in the control populations, up to 40% of the variance was accounted for in the regression models. This means that the effects shown were either stochastic, or by chance. It could be that the relationships predicting such effects were largely non-linear, or there were strong interaction effects between the predictors. Comparing the three intervention populations, the compensatory group appeared to be predicted by non-pathology factors. In general, this population showed a higher learning rate and higher general ability, and so successful intervention effects could have been the product of some sort of cognitive reserve. In contrast with this, normalisation populations were, in general, predicted by the pruning pathology.

5.8 Discussion

The aim of this chapter was to explore intervention effects in atypical populations, with the intention of comparing the differences between subgroups within the populations, the extent to which timing of intervention affected overall treatment effects, and to understand the mechanistic basis underlying individual variability. Overall, positive effects of intervention were produced across the three early intervention populations. A total of 14 populations showed some positive intervention effects (see Table 5.1). However, in all cases, such effects were limited. I discuss the results and suggestions for further intervention work below.

The role of timing explained some of the variability of intervention response, with earlier intervention populations demonstrating larger intervention effects in comparison with the compensatory late population. Clinical findings have also demonstrated significant effects of intervention in ASD when initiated in infancy. One example is the Infant Start Intervention (Rogers et al., 2014), which tested seven infants aged between 6 and 15-months that

exhibited a number of target symptoms of ASD (e.g. abnormal repetitive behaviours). By the age of three, the children demonstrated a reduction in autistic like symptoms in comparison with children whose parents had declined intervention.

Results from the two measurement points; end of intervention and end of training showed that in both the compensatory and normalisation early populations, intervention effects persisted until the end of training. Figure 5.3 illustrates this point, showing the mean intervention effects across measurement type and population. This is particularly interesting because of the comparisons that can be made with clinical intervention data. For example, a prospective follow-up study of children that participated in the Early Start Denver Model (see introduction) showed that on average, children maintained the gains from early intervention two years after intervention had ceased (Estes et al., 2015).

Combinations of intervention type and timing produced varying treatment outcomes. Early compensatory populations produced the stable effects of intervention, producing positive effects both at the end of training (70 epochs) and the end of intervention (1000 epochs) when either Regular or Rule measurements were implemented. Compensatory groups also demonstrated significantly higher treatment-effect scores at the end of training across Regular and Rule measurements in comparison with equivalent timing and measurement conditions in early populations. An improvement in overall effect scores persisted for the remainder of the training set; therefore demonstrating that short periods of additional experience with a particular set of behaviours could produce small long-term changes in recovery if intervention was provided early in training. A higher pre-intervention ability level significantly predicted positive intervention effects in the compensatory populations. This finding complements clinical data reporting associations between higher IQ and positive

outcomes later in development (Lord et al., 2014; Fein et al., 2013). Individuals in the normalisation early population were positively affected by intervention measures across all measurements. In contrast to the compensatory groups, however, greater intervention effects were demonstrated across the irregular measures. Results from the normalisation category suggest that in general, intervention effects are greater by the end of training, rather than at the end of intervention. One prediction for a normalisation population with a late intervention would be that the later the intervention starts, the less successful normalisation training set would be, and the more important compensatory effects will be.

Chapter 4 addressed the question of variability in outcome of ASD. The subgroups identified were utilised here in order to explore variability in long-term recovery and investigate whether differences between groups and response to intervention would be identified. Significant effects of intervention were identified in the Lower, Slower and Typical subgroups across all populations and at both end of training and end of intervention. Overall, Messy and Lower sub-groups were least affected by intervention across all three populations. Conversely, Typical networks exhibited the highest treatment effects in the compensatory groups, and the Slower group in the normalisation category. If we were to compare the Slower subgroup to a 'delay' group, an interesting suggestion from the results is that intervention appears to be more successful under circumstances of delay, rather than marked atypicality. The distinction between resolving and persisting delay, and the neurocomputational factors that predict the difference is considered by Thomas and Knowland (2014) in a similar model.

Multiple regression analyses were run in order to look at the individual response to intervention. All populations and subgroups that demonstrated significant intervention

effects were included for analysis. A total of nine were identified. A striking comparison can be made between the normalisation and compensatory populations here; the results suggested that predictors of variability depend on intervention type. The compensatory results showed that in general, individuated response depended on properties related to cognitive reserve, such as those indexed by IQ. Conversely, normalisation results showed that individuated response was modulated by differences in the severity of the pathology related parameters. A key aim of using computational models is to advance developmental theory and understand mechanistic properties underlying cognitive functioning. This finding demonstrates a novel, testable prediction that has not yet been proposed in modelling or clinical interventions.

Overall effects of intervention were very small in comparison with the scale of behavioural deficits in the HREO population. Low effect sizes in the regression models demonstrate the challenge of understanding the relationship between the variability in the model and the possible predictors involved in the outcome of each population. The largest amount of variability accounted for in the model was 40%, and it is likely that higher order interactions between parameters are accounting for much of the unexplained variance, somewhat limiting the generalisability of main effects reported here. That being said, beneficial effects of intervention were confirmed across the population groups. Furthermore, intervention persisted until the end of training for some early intervention groups. Parameters predicting effects of intervention were also identified, with the finding that compensatory and normalisation group effects were modulated by different parameter types. Findings from the model were also consistent with hypotheses from clinical intervention trials suggesting that earlier intervention can lead to positive outcomes in ASD (e.g. Green et al., 2015; Kasari et al., 2014; Landa & Kalb, 2012). See Chapter 1 for an overview of early intervention).

In terms of small effect sizes, Yang and Thomas (2014) argue that limitations in the computational properties in associative systems may be hard to overcome simply by altering the training set of the networks. One possibility would be to allow the intervention set to be permanently embedded into the normal training set. It could be argued that the intervention set will enter the broader conceptual knowledge of the child's system, which is likely to be refreshed by interaction with the episodic memory system.

The settings that were used in the intervention analyses were preliminary. I believe it would be of use to run interventions on the three populations at an earlier time point, as this analysis has determined that earlier interventions produce a more beneficial treatment effect. It will also be of interest to compare a normalisation late population with the compensatory late category to see whether there would be a difference in individual variability. One finding from the results in the current chapter was that variability in the compensatory early group was modulated by pre-intervention ability, whereas normalisation populations were affected by pathology-based parameters. It could therefore be suggested that those who have a greater impairment pre-intervention are more likely to recover some ability if they are in a Compensatory population. It would be interesting to explore this idea by comparing intervention effects on highly damaged systems in compensatory and normalisation intervention types.

Chapter 6

6.1 Introduction

In this thesis, two central questions were addressed. The first concerned variability observed in the phenotypic behavioural profiles of infants with ASD, and our ability to predict variability and clinical outcomes based on cognitive and observational measures in infancy. The second concerned the identification of homogenous subgroups in ASD and underlying mechanisms that distinguish them. Data from a longitudinal prospective study of high-risk infants and a developmental computational model of ASD were combined to explore the influences shaping developmental trajectories. Here, it is argued that in order to elucidate the mechanistic influences underlying ASD, it must be established whether putative homogenous subgroups and causal pathways can be identified. This discussion chapter will begin by summarising the key aims and subsequent findings of the thesis. I then discuss the implications of the results and position findings from this thesis in relation to relevant research within the field of ASD. Finally, I discuss methodological and theoretical limitations and consider specific study modifications and directions for future work.

6.2 Aims of the thesis

The primary aims of this thesis were as follows: 1) to combine the analyses of data from a prospective, longitudinal dataset and a computational model of ASD to examine the influences of variability across infancy. With respect to the prospective empirical data: 2) to compare within- and between-subject variability in cognitive and behavioural measures

between high-risk and low-risk infants. 3) To elucidate the extent to which cognitive and environmental factors predict outcome in high-risk populations. With respect to the modelling: 4) to clarify the influences on developmental trajectories by identifying regressive and non-regressive subgroups in the computational simulations and to examine the underlying mechanistic differences in each of the homogenous subgroups. 5) To implement interventions in the simulated high-risk population and assess the effect of timing and intervention type on developmental outcome.

This is the first set of studies to utilise data from both a prospective, longitudinal dataset and a computational model to assess the predictive power of subgroups and variability in ASD. The majority of computational models looking at ASD have focused on single deficits within the disorder (e.g. McClelland, 2000; Beversdorf, Narayanan, Hillier & Hughes, 2007). Furthermore, Over-Pruning model is the only account that has considered the variability and the identification of risk factors. Computational models can be used in conjunction with clinical studies of ASD, and here they are used to clarify theoretical perspectives, narrow the set of questions that should be addressed and produce novel, testable predictions. Whilst the model employed here relied on abstract training sets rather than targeting the emergence of a specific behaviour, it nevertheless was able to serve as a complementary tool to understand the role of causal mechanisms in ASD. I used clinical data to identify the extent to which cognition, environmental factors and intra- and inter-subject variability could be predictive of ASD outcome. Findings were related to predictions from the Over-Pruning hypothesis that emerged from the computational account (Thomas, Davis Karmiloff-Smith, Knowland and Charman, 2015). Computational simulations were then utilised to identify more homogenous subgroups within simulated individuals and consider the underlying neurocomputational mechanisms that distinguish each subgroup. Simulations were also used to conceptualise the

effects of intervention based on the timing of onset, duration, and intervention type. This is an innovative method in the field of computational modelling in ASD.

The investigation of clinical data was presented in Chapters 2 and 3. Chapter 2 investigated intra- and inter-subject variability in high-risk infants from the BASIS prospective Phase 1 cohort. Variability has been explored in terms of stability of diagnoses over time and subgroupings based on cognitive and behavioural phenotypes. However, variability has not been utilised as a measure that could potentially distinguish typically developing high- and low-risk individuals from infants who are later diagnosed with ASD. I showed that high-risk infants could be distinguished by inter-subject variability scores in developmental measures, but only at 36-months. At this point, variability in motor, language and communicative domains was significantly higher in high-risk infants with an ASD diagnosis in comparison with high-risk non-ASD and low-risk infants. A single variability score computed at each time point across cognitive measures represented an uneven cognitive profile. A more uneven cognitive profile (defined as demonstrating strengths in some cognitive domains, and lower ability levels in others) at 24 months was associated with poorer expressive and receptive language, and social abilities at 36 months in high-risk infants. However, high-risk ASD and non-ASD infants were not significantly different, suggesting that unaffected siblings may exhibit milder versions of autistic traits, and therefore similar levels of heterogeneity, which is congruent with the idea of the broader autism phenotype (Ozonoff et al., 2014). A more uneven cognitive profile was not associated with diagnostic outcome at 36 months. A significant increase in a second intra-subject variability measure (between time points across single domains) was found in high-risk infants who were later diagnosed with ASD. Between 24 and 36 months, variability in expressive language significantly increased,

whereas in visual reception it significantly decreased. Neither score was an independent predictor of ability at 36 months.

Chapter 3 focused on predictors of later outcome in prospective studies of ASD. Numerous theories focusing on the early manifestation of ASD propose that the primary emerging symptoms will be social in origin, and that non-social symptoms will be exhibited later in development (Chevallier, Kohls, Troiani, Brodkin & Schultz, 2012; Kawakubo, et al., 2007). However, the current consensus from prospective research is that autistic symptoms appear simultaneously in both social and non-social domains (Jones et al., 2014). Furthermore, the Over-Pruning account predicted that primary deficits would be identified first in sensory and motor areas, and that atypical predictive behaviours would affect multiple cognitive domains (Thomas, Davis, Karmiloff-Smith, Knowland & Charman, 2015). I assessed the predictive power of a number of early risk markers across social and non-social domains to clarify these inconsistencies. Here scores from social, communicative, motor and environmental measures at 7 and 14 months were used as predictive measures of developmental outcome at 24 months and 36 months. Neither social nor non-social measures from 7 months were predictive of diagnostic outcome. This was consistent with other prospective findings in the ASD literature (Jones et al., 2014). However, by 14 months a combination of social, motor and environmental factors predicted outcome with 77% accuracy. Specifically, lower expressive language abilities and a lower environmental score based on socio-economic status measures (specifically, a lower household income and father's income based on job type and father's highest qualification) increased the likelihood of being categorised as above the threshold for ASD (based on the ADOS diagnostic instrument) at 24 months. A lower fine motor and higher AOSI score, whilst not significant predictors in isolation, increased the reliability of the regression model. This finding indicated that higher-order factors were an influence in

the statistical model. The validity of the model was confirmed by adding coefficients from the predictive regression model into a second model that utilised data from a second, independent cohort of BASIS infants, known as Phase 2. The new model correctly predicted 81.5% of infants above the threshold for autism (based on the ADOS diagnostic instrument and not clinical diagnosis, due to the age of the preliminary outcome and data available) at 24 months. Sensitivity was 71.4% and specificity was 81.5%. The positive predictive value was 70.5%, and negative prediction value was 96%. When infants above the threshold for ASD and autism were combined (individuals displaying severe and mild symptoms at 24 months) the total accuracy rating was 77%, sensitivity was 78.5% and specificity was 81.4%. However, the positive prediction value was 55.5%, whilst the negative prediction value was 91.6%, demonstrating a higher error rate when categorising infants to be above the threshold for ASD and autism. Therefore, the model was more likely to incorrectly categorise individuals (as above the diagnostic threshold) when the sample included infants that exhibited less severe symptoms. Whilst a low sample size placed limits on the generalisability of findings, this study was the first to directly validate model predictions from one prospective cohort using infants from a second, independent cohort where developmental outcomes were not known prior to the generation of predicted outcomes from statistical model.

Turning to the computational modelling chapters, Chapter 4 analysed three simulated populations from an artificial neural network model of ASD, with the aim of identifying regressive and non-regressive subgroups based on developmental trajectories. Here, I showed that regression was significantly more likely to occur in simulations at high-risk of pruning compared to a control population. I then considered simulations in the high-risk early onset population where both the pruning threshold was set to a higher level, increasing

the risk of aggressive pruning, and an earlier onset of pruning was more likely. This population displayed lower mean scores compared to simulated individuals in the original high-risk population. This pointed to an interaction between pruning onset and threshold that led to adverse outcomes. Simulated individuals not showing overt behavioural signs of regression were split into high and low performance groups based on final outcome scores. As a group, networks that exhibited higher outcome scores had lower pruning thresholds and a lower likelihood of pruning combined with a higher learning capacity. Non-regressive subgroups were then identified based on the developmental trajectories of each simulation. Statistical regression analyses were then run to identify the mechanistic differences underlying each subgroup. One typical (Typical) and three atypical groups (Messy, Lower and Slower) were identified, separate to the regressive subgroup. Typical networks demonstrated consistent development throughout training, Messy networks exhibited unstable responses that oscillated throughout development, Lower networks exhibited poorer development over time, with final accuracy levels at 60% or below, and Slower networks demonstrates slower development, but higher levels of accuracy than the Lower networks (see Chapter 4). Typical groups were distinguished by a non-pathological pruning mechanism and higher learning rates, whilst a lower learning rate was one of the largest predictors that distinguished atypical groups with pathological pruning. In the same population, I compared recovery rates in simulations displaying regression. Rates of recovery improved as learning rate and capacity increased and as the pruning threshold decreased, suggesting that both variation in the severity of pathology as well as population-wide individual differences could be responsible for rates of development in affected individuals. Regressive networks were split into high and low groups based on final outcome scores. Analyses comparing the two categories showed that a better outcome was based on lower pruning thresholds and a later onset of pruning. Overall, the model provided evidence

that a single pathology (of varying severity) could produce diverse deficits and trajectories of development through interactions with population-wide individuals differences. A range of non-regressive and regressive subgroups was produced by a single underlying common cause, over-pruning. However, individual differences in neurocomputational properties were identified in subgroups in the population, creating risk and protective factors across multiple systems. In these simulations, either the pruning pathology was milder or occurred when development had a greater opportunity to strengthen connectivity to be more robust against pruning.

Chapter 5 utilised networks from the high-risk early onset population to investigate the effects of intervention. It is argued that developmental computational models can be used as a complementary tool to clarify theories from clinical research, and to elucidate the mechanistic processes that underlie the response of complex developmental disorders to behavioural intervention (Mareschal & Thomas, 2007). Here two intervention types were implemented. The first aimed to normalise the atypical system, where normalisation was defined as bringing performance into the typical range for the full range of behaviours characterising the target cognitive domain. The second attempted to optimise performance on regular verb learning, a subset of the full training problem, which was defined as compensation in the learning system. Intervention effects were then split by the subgroups identified in Chapter 4. The timing of intervention was considered by contrasting early and late intervention points for a fixed period. Compensatory and normalisation intervention types both demonstrated small, but significant positive intervention effects overall. In terms of timing onset, effects were reliably larger when intervention was implemented an earlier time point. Interestingly, networks in both compensatory and normalisation interventions displayed positive intervention effects both at the point that intervention ceased and at the

end of training, which suggested a long-term impact from both intervention types. However, simulations in the compensatory intervention group demonstrated positive intervention effects in regular and rule generalisation tasks, whereas gains were identified in irregular word learning tasks in normalisation intervention set. The neurocomputational parameters that modulated each intervention type were identified. Whilst successful compensatory intervention relied on cognitive reserve such as pre-pruning ability levels, normalisation effects were modulated by the severity of the causal pruning mechanism. Finally, in terms of subgroupings, intervention demonstrated some success in Typical, Lower and Slower networks. Messy and Lower networks benefited least, because of their lower learning capacities and earlier pruning onset. A high level of variability was identified in response to intervention across networks, a finding which is also reported in clinical findings (Green et al., 2015; Dawson et al., 2012). The model, which addressed disorder and variability, is uniquely placed to investigate this finding.

6.3 Implications for clinical research

The research in this thesis presents a number of implications that are congruent with current findings in prospective studies. Chapter 3 presented a predictive model of ASD where environmental factors and developmental ability scores at 14 months were predictive of diagnostic outcome at both 24 months and 36 months. At 7 months however, development assessed by Mullen, Vineland and AOSI scores and SES measures were seemingly typical for high-risk infants who were later diagnosed with ASD. Low-risk and high-risk infants were indistinguishable in social, communicative or sensory and motor domains at this point. It is consistent with the majority of findings from prospective, high-risk studies, which have yet to demonstrate substantial evidence for early behavioural markers at 6 months (Jones et al., 2014; Zwaigenbaum et al., 2013). This is also consistent with hypotheses from the Over-

Pruning account (Thomas, et al., 2015). Specifically, it was proposed that autistic symptoms would not emerge until synaptic pruning has commenced. Pruning occurs at different times in different regions of the brain following a stage of initial overgrowth in the first few months of infancy (Huttenlocher at al., 1997). Therefore, prior to pruning, the Over-Pruning account predicts a phase of seemingly typical development. Looking to neurobiological findings, Uddin, Supekar and Menon (2013) propose that over-connectivity could be seen as a feature of early development, whilst under-connectivity is identified in later development in children who go on to develop ASD. This is supported by findings from Tager-Flusberg and Nelson (2013), who used near-infrared spectroscopy to measure connectivity across early development in infants at a higher risk of developing ASD. They reported that increased connectivity was identified at 3 months and decreased functioning connectivity compared to low-risk controls at 12 months.

A second hypothesis from the Over-Pruning account was that motor domains should be the first to exhibit atypicalities. It is recognised that typical synaptic pruning occurs initially in low-sensory and motor areas (Gotgay et al., 2004) followed by higher association areas (involved in behaviours such as speech and abstract thought) and last in the prefrontal cortex (Huttenlocher & Dabholkar, 1997). Therefore, if pruning is atypical, then sensory and motor domains will be the first to exhibit atypicalities. However, the current findings were not consistent with this theory. Here, individuals who were later diagnosed with ASD showed the earliest atypicalities in fine motor skills at 12 months, but simultaneously in expressive language and AOSI scores (measuring levels of autistic traits). One explanation for the lack of differences identified between 6 and 12 months in motor domains is that Mullen gross and fine motor measures are relatively insensitive to subtle differences in behaviour. Thus, it is possible that they may not provide enough sensitivity to identify early, subtle motor

differences in young infants. Some studies using standardised tests of fine and gross motor abilities have failed to identify differences before 12 months when comparing high-risk ASD and low-risk groups (e.g. Ozonoff et al., 2010; Bhat, Galloway & Landa, 2012; Focaroli, Taffoni, Parsons, Keller & Iverson, 2016). However, Estes et al. (2015) reported less advanced gross motor skills at 6 months in high-risk infants who were later diagnosed with ASD. They propose that sensorimotor differences precede other cognitive atypicalities and behavioural features of ASD between 6 and 24 months. Furthermore, a small number of prospective studies focusing on subtle motor deviations have supported the Over-Pruning. Flanagan, Landa, Bhat and Bauman (2012) found that infants who were later diagnosed with ASD displayed more frequent head lag and less postural control than low-risk or high-risk infants. Zwaigenbaum et al. (2005) showed that parents reported lower activity levels in infants at 6 months who were later diagnosed with ASD, with a possibility that atypical posture control could be a contributing factor. Taken together, these studies highlight a potential early motor atypicality that could be utilised as a marker for diagnostic and intervention referral.

Three points can be made in relation to these findings. First, it is possible that early atypicalities are present in motor domains, but are only identifiable when investigating fine-grained motor abilities. Alternatively, due to the challenges of assessing subtle atypicalities in young infants, it is possible that differences are in part due to high error rates (Jones et al., 2014). A third consideration is that early motor atypicalities identified in ASD are indistinguishable from infants that demonstrate developmental delay and are therefore not specific to ASD. For example, head lag is present in infancy across other disorders such as cerebral palsy (Barbosa et al., 2005), but it is not currently understood whether the causal mechanisms behind comparable phenotypic atypicalities are the same across disorders. That

is, because early motor atypicalities are found in other developing disorders does not in itself alter the possibility that these are a core early emerging feature of ASD. Whilst the issue of defining early syndrome-specific markers remains unresolved, it is perhaps an indication that future research, where possible should look to include more sensitive sensory and motor measures to elucidate current claims.

In Chapter 3, no single behavioural measure was identified as an independent predictor of outcome. Instead, the combination of multiple measures significantly predicted developmental outcome. Lower levels of fine motor and expressive language abilities, combined with a higher AOSI scores were predictive of ASD outcomes at 24 months. Two findings are important here. First, this finding contradicts research proposing that social deficits will be first to appear in infancy. As discussed in Chapter 1, a number of prominent causal theories of ASD focus on social processes such as attention that contribute to the overall development of social skills (Landry & Bryson, 2004; Mundy & Neal, 2001). Such theories propose that primary deficits in precursor social domains cause downstream atypicalities in secondary skills, which limit social and communication skills and in turn produce restricted and repetitive interests and the full phenotypic manifestation of ASD symptoms.

To date, two prospective studies have provided evidence for the presence of early social atypicalities at 6 months. Chawarksa Macari and Shic (2013) reported a reduction in gaze fixation to humans in social scenes at 6 months in infants who were later diagnosed with ASD. Jones and Klin (2013) reported decreased fixation to humans in social scenes, specifically in the eye region, between 2 and 6 months in comparison to low-risk infants.

However, looking at data from 6 months, the results are contradictory. Jones and Klin failed to identify differences in social orienteering at 6 months (as reported by Chawarska et al., 2013). Moreover, a reliable group difference between children who were later diagnosed with ASD and typically developing controls was not identified by Jones and colleagues until 12 months. The contradictory results are enough to suggest that both findings are preliminary. In order to consider either as significant early social predictors of ASD, robust replications will be necessary. Furthermore, interventions focusing on social domains have not demonstrated widespread improvements by focusing on single atypicalities. In a social account of ASD it might be assumed that interventions focusing on early social deficits should provide an opportunity to improve other areas of cognitive functioning. By alleviating primary deficits, so too should secondary consequences be alleviated. Thus far, this has not been the case. Kaale et al. (2014) presented data from a longitudinal follow up intervention study in which half of the infants received additional joint attention intervention as well as in general language and cognitive domains. Whilst both groups displayed gains in language and cognitive outcomes and the joint attention group also demonstrated gains in joint attention, no other significant group differences were observed. Additional joint attention based intervention did not produce additional improvements in general areas of language, cognition or other social domains. Taken together, current findings from prospective studies, together with the results presented here provide little evidence for social deficits preceding atypicalities in other domains (Jones et al., 2014; Zwaigenbaum et al., 2013).

A second finding here is that multiple measures predicted outcome rather than a single deficit in one domain. This poses a problem for accounts predicting that single deficits in one domain will produce secondary deficits in ASD (for example, the social brain theory

(Hadjikhani, Joseph, Snyder & Tager-Flusberg, 2006), the Theory of Mind account (Baron Cohen, Leslie & Frith, 1986), and the Weak Central Coherence theory (Happe & Frith, 2006), as well as theories of social orienting and attention (e.g. Dawson, Meltzoff, Osterling & Rinaldi, 1998; Bryson et al., 2004). The implications of this finding lead to a number of possible conclusions. The first is that a single causal pathway could produce the co-occurrence of social and non-social behaviours. The Over-Pruning account offers one potential account of observed difficulties. Since pruning is a widespread neural mechanism, it is predicted that atypicalities will be widespread and time-varying, with atypicalities in structural brain connectivity emerging through development. These will be expressed as deficits and compensatory mechanisms based on experience-dependent individual differences, (Thomas et al., 2015). An alternative account is the Interactive Specialisation theory (Johnson, 2011), which also accounts for multiple deficits across ASD. Here, the suggestion is that atypical processing affects early infant experiences, which over time can cause negative downstream consequences, which in turn, strengthens atypical developmental trajectories. Alternatively, multiple predictive measures could be explained by separate neural mechanisms. For example, Happé and Ronald (2008) (see also Happé et al., 2006) utilised a genetically sensitive design involving a cohort of 3000 twin pairs to demonstrate that social and non-social behaviours in ASD demonstrated only modest correlations, and whilst both trait types were highly heritable, there was only a small genetic overlap. In this study however, autism was treated as a trait that varied across the whole population, and participating children were placed on a scale of severity for each trait. Thus some caution must be taken when comparing genetic study designs with prospective ASD data.

There is an increasing consensus that intervention should target a wide range of cognitive domains. The most successful intervention effects thus far have included both social and

non-social cognitive and behavioural measures (Dawson et al., 2010; Green et al., 2010). In a review of intervention success thus far, Zwaigenbaum et al. (2014) suggested that intervention should combine tasks from a range of cognitive and behavioural domains to optimise developmental outcomes. The need for wide-ranging intervention fits less intuitively with a narrow originating deficit. Looking to hypothesis from the Over-Pruning account, it is argued that intervention will only provide successful changes in the specific system that it is targeting. In ASD, atypicalities will be system-wide due to the nature of over-aggressive synaptic pruning. Therefore, intervention must target a broad range of behaviours in order to compensate across multiple atypical neural systems, and to the extent that pruning differentially impairs long-range connectivity, the interventions must employ behaviours that simulate such connectivity.

An environmental measure consisting of household income, father's job type and highest educational level was also a significant predictor of outcome in infants who were later diagnosed with ASD. Overall, a lower environmental measure was associated with an increased likelihood of having an ASD diagnosis at 24 or 36 months. Conclusions based on this measure are cautious here. Looking to findings in the current literature, results are relatively mixed in terms of the impact of SES factors. Studies have reported a link to low SES and poorer developmental outcomes, and a lower success rate after intervention (Bradley & Corwyn, 2004; Carr et al., 2015). Furthermore, Fein et al. (2013) suggest that optimal outcome may be less prominent in individuals with a lower SES. Conversely, Larsson et al. (2005) found that SES was not a contributing factor in ASD outcome severity. Of course, the reasons why a lower SES may be predictive of ASD outcome in at-risk children are difficult to disentangle. One possibility is that infants from a higher SES may experience a more stimulating environment. Looking from the perspective of the Over-Pruning account,

increased environmental stimulation in early infancy could produce protective factors, which could potentially strengthen particular cognitive abilities and protect somewhat from pruning mechanisms. Alternatively, in later development, infants with a higher SES may be more likely to experience more optimal interventions, and fewer family stressors (Fountain et al., 2012). Before environmental effects are confirmed here, further consideration of the measures that were utilised will be necessary. In future studies it would be valuable to consider specific parental and other environmental measures in order to extricate precise predictive measures which are often overlooked in prospective research, given that so many neurocognitive factors are confounded with SES (Hackman & Farah, 2009, 2010).

In this thesis, two further logistic regression models were run using infant behavioural data in order to validate the generalisability of findings from the preliminary predictive model. Both models used a combination of data from the Phase 2 BASIS cohort with coefficients from the significant regression model from Phase 1 to predict Phase 2 outcomes. The first correctly predicted 77% of individuals as being above the threshold for autism. In the second model, when infants above the threshold for ASD (demonstrating a milder impairments overall) and autism, were included, the accuracy level dropped to 55%. This was accounted for by false positives, that is, the model incorrectly classified infants as above the threshold for ASD and autism. As discussed in Chapter 3, a low sample size, particularly for individuals in the 'above autism threshold' (n = 7) makes generalisability of these findings uncertain. However, these findings identify a subset of infants who displayed more severe ASD symptoms at 24 months on the basis of cognitive atypicalities and environmental factors at 14 months in comparison to individuals who exhibited milder ASD symptoms at 24 months. Thus far, a number of prospective studies have identified putative subtypes of autism based on deviations in developmental trajectories within the first three years of infancy. Landa and

Garrett-Mayer (2006) initially identified two autistic subgroups. The first, known as 'early onset', displayed impairments in social and communicative domains by 14 months, whilst other infants, known as the 'late onset' subgroup, did not demonstrate ASD symptoms until the second or third year of infancy. Landa et al. (2007) also showed between-group differences at 36 months. Infants with signs of ASD at 14 and 36 months had fewer consonants in syllables compared to other high-and low-risk outcome groups. However, this difference was not found in infants who did not display ASD symptoms at 14 months but were diagnosed at 36 months. Others (Ozonoff, Heung, Byrd, Hansen & Hertz-Picciotto, 2008; Ozonoff et al., 2011) have identified more subgroups based not only on the onset of symptoms, but also to include the developmental trajectory sub-groupings of regression, plateau and a number of individuals who display a combination of categories. If subtle behavioural differences between ASD subgroups could be identified, it provides a possibility for carefully targeted interventions that could focus on differential areas of weakness.

Looking at the results from chapter 3, the statistical regression model significantly predicted diagnostic outcome with reasonably good sensitivity and specificity levels. Furthermore, this was corroborated by an independent dataset, providing validation for the utilisation of these predictors. However, a proportion of infants were not correctly classified as being above or below the threshold for ASD. One possibility for this finding is that a subgroup of individuals with ASD are characterised by a lower SES, lower motor skills, lower expressive language and higher AOSI scores. Whilst the statistical regression model was successfully verified using a second dataset, testing this profile on larger groups of individuals who are later diagnosed with ASD would be necessary to confirm this finding.

Looking to the mechanisms that may underlie heterogeneity seen in behavioural manifestations of the disorder, I discuss findings from novel modelling populations that I implemented in Chapter 4 (taking the view of the Over-Pruning account). I extended findings from the original Over-Pruning model (Thomas, Knowland & Karmiloff-Smith, 2011), which sought to examine regression in ASD. Here I showed that the one model could explain multiple subtypes of ASD due to the pathological pruning mechanisms interacting with individual differences in other neurocomputational parameters, such as the onset of pruning, learning rate and capacity. Whilst the pathological process is system-wide, individual differences in the form of risk and protective factors may be differential across systems. Thus, in high-risk infants, some individuals may exhibit a more robust prefrontal system, or sensory system for example, which could account for differences in cognitive profiles in ASD. Thomas et al. (2015) argue that this prediction is testable if one assumes that risk and protective factors are themselves partially heritable, and the variation seen in cognitive profiles in ASD therefore be assessed by testing for the presence of (typical) consistent profiles of cognitive strengths and weaknesses in typically developing family members.

In Chapter 2, I examined intra-subject variability in cognitive domains between 6 and 36 months. One question assessed the extent to which an uneven cognitive profile was predictive of later diagnostic outcomes in low-risk and high-risk infants. At 6 months, low- and high-risk infants displayed similar levels of variability across cognitive domains. This is again in line with the idea that the first few months of infancy appear to indistinguishable in infants who are later diagnosed with ASD. Interestingly, divergence in heterogeneity between low- and high-risk individuals occurred at 24 months and an uneven cognitive profile was negatively correlated with visual reception and language scores at 36 months.

However, high-risk non-ASD infants also exhibited uneven profiles. This supports the idea of the broader autism phenotype (Ozonoff et al., 2010; 2014), that is an uneven cognitive profile may be present in high-risk non-ASD individuals because they display traits that are common in ASD, but not across all domains. The Over-Pruning account also predicts high levels of heterogeneity in unaffected high-risk infants. It argues that infants may be exhibiting milder versions of the pruning pathology, which also interacts with individual differences in the form of risk and protective factors, leading to fluctuating profiles in high-risk individuals. Alternatively, an uneven cognitive profile could be a non-pathological risk factor that makes the effect of atypical pruning worse, perhaps representing less well functionally connected systems that are more prone to disruption by structural connectivity pathology.

6.4 Implications from computational modelling

Caution must be taken when drawing inferences from computational models. In this iteration of the model, relatively simplistic and abstract learning tasks were used as measures of development (albeit comparable to the complexity of other existing models of ASD). However, the model was conceived using available data on behavioural regression, as well as genetic and biological findings, and was based around a strong hypothesis-driven theory. Comparing the fit of computational results with clinical findings has not only accelerated the determination of the hypotheses proposed in the Over-Pruning account, but can also inform clinical data as to directions for further research.

In a previous iteration of the Over-Pruning account (Thomas et al., 2015), we proposed that intervention would not be able to restore connectivity that had already been lost through

over-aggressive pruning. The findings in Chapter 5 provide a more solid mechanistic foundation to this claim. Whilst positive intervention effects were identified in both normalisation and compensatory interventions, total recovery was not apparent in atypical populations. Furthermore, intervention implemented at an earlier stage of development was more effective in both compensatory and normalisation conditions. These findings would suggest that it may be possible to maximise abilities that are already present by utilising the connectivity that remains, tailored to the types of computation which connectivity can support. When intervention was implemented earlier, it was effective in strengthening the remaining connections against further damage from pruning. Development prior to 2 years of age is a time of accelerated connectivity, brain growth (Wolff et al., 2012) and neural plasticity (Mundy & Neal, 2000). If interventions could be employed in infants before this process is reduced (and as is predicted here, prior to the full effects of pruning), it may be possible to modify cortical organisation and improve the trajectory of development by enhancing learning (Fox et al., 2010; Johnson, 2001; Webb, Kelly & Dawson, 2014). Thus far, the small numbers of clinical studies that have implemented early intervention have demonstrated preliminary beneficial effects (Bradshaw, Mossman Steiner, Gengoux & Koegel 2015) using early intense behavioural interventions (EIBI; Howlin, Maglatti & Charman, 2009) or those that combine behavioural, and social and communicative approaches (Early Denver Start Model (EDSM): Dawson et al., 2010, 2012; Rogers et al., 2012).

A second comparison from the model that could further support this theory would be to compare late compensatory and normalisation intervention sets. The hypothesis here would be that the later intervention takes place, the more likely that a network would respond to compensatory training mechanisms to focus on utilising remaining connectivity, rather than

attempting to normalise a system that no longer has the resources to support the target function. Further simulations could test this proposal more directly. Therefore, later compensatory intervention effects may be more beneficial than normalisation attempts. At the time of writing, late normalisation training data were not available for analysis, but it is a consideration for further research.

In Chapter 5, I reported the finding that a simulation of a short burst of early additional training akin to a behavioural therapy, led to improvements that while modest, lasted not only to the end of intervention, but also to the end of training. Estes et al.'s study (2015) also supported the possibility of long-term effects. Children were assessed two years after partaking in the EDSM of early intervention. On average, gains from the intervention were maintained two years later, including the reduction of core ASD symptoms in comparison with a community intervention group. The majority of data have shown positive short-term intervention effects (Green et al., 2010; Posalawsky et al., 2014; Kasari et al., 2014). However this is in part due to the lack of follow-up studies to date.

Clinical intervention effects in ASD are varied, and whilst some studies have demonstrated positive effects in a number of individuals, there are also infants who fail to make developmental gains (e.g., Green et al., 2015; see Chapter 1). I considered the mechanisms underlying individual differences and differing responses to intervention in the simulation populations. In all populations in the model, equivalent measures for IQ, that is pre-intervention learning ability and general ability (which was derived independent of environment and in the absence of the pruning pathology) were both predictive of outcome. A higher pre-intervention ability level was predictive of a more successful post-intervention

outcome. This finding is consistent with some clinical studies that have reported pre-intervention IQ scores as a predictor of outcome at the end of intervention (Howlin, Magiati & Charman 2009; Smith et al., 2010). However, Zachor, Ben-Itzchak, Rabinovich and Lahat (2007) concluded that whilst children with higher IQ scores displayed higher social and language skills after intervention training, children with a higher IQ did not make significantly larger developmental gains than children with a lower IQ scores.

When considering underlying mechanistic differences at a group level, however, the issue of variability remains problematic. In Chapter 4, I successfully identified subgroups based on developmental trajectories. As well as assessing the individual differences that led to different responses in each subgroup, I also assessed the computational parameters that accounted for group membership. As a group, typically developing simulations were characterised by a higher learning rate, a lower pruning rate and later onset of pruning. In contrast, a lower learning rate was indicative of a slower trajectory, and a low learning rate with a higher pruning probability generated poorer developmental outcomes. However, at the individual level, high levels of variability were present in terms of the effect of each computational parameter. So, whilst trajectories exhibited similarities to the point where they could be classified into specific subgroups, group parameter effects masked some of the within-group variability that accounted for individual differences in outcome. This is also true when identifying subgroups in clinical data. For example, individuals may demonstrate regressive or early or late onset patterns in development. This does not mean that these individuals will respond in the same way to interventions. I would argue that even if subgroups are identified, interventions must still be tailored to individual profiles to allow for targeted intervention based on abilities. By focusing on areas of cognitive strengths, targeted abilities could be nurtured and more positive gains may be possible.

6.5 Limitations and future directions

One of the main limitations of the clinical work presented in this thesis is that all data for individuals who went on to have an ASD outcome were collected from multiplex families (those with multiple children or family members with an ASD diagnosis). The generalisability of the findings is hampered when data consider only multiplex infants, as one must assume that the manifestation of ASD in multiplex facilities is also representative of ASD occurring in simplex families (where ASD is not present in siblings or first degree family members). Indeed, a number of studies have highlighted behavioural differences between simplex and multiplex families in the pattern of social impairments displayed (Constantino, 2002). Some studies have reported higher severity levels of social impairments in individuals from multiplex families (Virkud, Todd, Abbacchi, Zhang & Constantino, 2009), whereas others have reported more severe ASD symptoms measured by ADOS-G in simplex probands (Taylor et al., 2014). One specific direction for a future study would be to utilise data from a simplex sample in order to investigate the divergence between simplex and multiplex families. An example cohort that could be utilised is Simons Simplex collection, where more than 2000 individuals with simplex ASD have been evaluated (Fischbach & Lord, 2010). Thus far, it has not been possible to create cohorts of simplex infants prospectively compared to multiplex families in at-risk studies. However, in this sample, standardised measures consistent with those utilised in the BASIS cohorts were used to assess individuals from 3 years. Whilst this database is not directly comparable to infants at the earliest time points, follow-up data from infants in the Phase 1 BASIS cohort known as (BASIS-7) cohort will be available (Shephard et al., 2016). One possibility would be to compare outcomes between 3 and 7 years in multiplex and simplex families.

A second limitation from the clinical work presented in this thesis is the relatively modest sample size. The number of individuals in the high-risk ASD and low-risk populations combined was under 100 in Phase 1. In Chapter 3, Phase 2 data that included 115 high-risk and 25 low-risk infants were used to validate a predictive model using 24-month outcomes. However, in Chapter 2, Phase 1 36 month outcomes were analysed, which had not yet been collected in Phase 2. This meant that the number of high-risk infants with ASD was 17 (as only Phase 1 data were analysed), which inevitably compromised the generalisability of findings. One way of rectifying this in future work would be to combine data from Phase 2 of the BASIS, where 36-month outcomes have now been finalised. One consideration when incorporating the data is that infants are from separate cohorts and slightly different methodologies were used across some measures. Furthermore, infants were tested across a larger range of time points in Phase 2, and future analyses would have to account for such differences. One way of overcoming this discrepancy is to utilise more complex methods of analysis that allow for measurement errors between groups.

A further limitation is that data only extends across the first 3 years of infancy. Later time points will allow a view of the stability of diagnosis and currently, infants in the BASIS Phase 1 cohort are being re-assessed at 7 years as a follow up to 36-month outcomes (Shephard et al., 2016). I plan to utilise these data to consider the predictive nature of within- and between-subject variability demonstrated at 24 and 36 months for later developmental outcomes. For example, in Chapter 2, infants with ASD demonstrated higher variability than the high-risk and low-risk groups, but only from 36 months. It would therefore be beneficial to compare heterogeneity at later time points to assess whether variability at 36 months is associated with later differentiated outcomes. Additionally, if SES effects pertain to the availability of resources, the reduction of stress in the family environment and family support

(among other things) rather than as a risk factor in original aetiology, one might expect such effects to exaggerate over time. Therefore, the assessment of children at 7 years would be important to elucidate this hypothesis.

Turning to the behavioural measures, one might ask whether Vineland and Mullen were the most appropriate measures to examine variability. These developmental measures have a number of advantages. Vineland assessments can be used from infancy through to adulthood. Furthermore, both tests contain a broad range of subdomains that assess motor, social, and communicative and language abilities. Scores are also standardised across measures, which enables cross-domain comparisons. It is possible that test measures containing a narrower range of assessments, but with more fine-grained testing measures in particular domains could capture variability where developmental assessments such as the Mullen and Vineland have not. To understand the variability seen in findings from prospective studies, measurements are required that can retain sensitivity across development.

One ethical consideration that has arisen from this thesis is the use of the term 'normalisation' in the computational model of intervention in ASD in Chapter 5. The computational definition of normalisation used here originated from a paper by Thomas and Yang (2015), which defines normalisation in computational models of intervention as "the acquisition of abilities and knowledge that a typically developing system acquires through exposure to training". It is clear that the term normalisation is not transferrable to clinical data. For example, Milton, Mills and Pellicano (2012) propose that the notion of normalisation in clinical research relates to the implication that normalisation is a primary objective in ASD research; much of this work fails to take into account the views of

individuals with ASD, many of which do not support the idea of a "finding a cure". Whilst in the current computational model, we do not define normalisation as in the clinical sense; its use has highlighted the need for clarity in the use of terminology that could have significant ramifications for the ASD community.

6.6 Computational limitations and future directions

One of the main drawbacks of utilising a computational model of development is the simplification of the general hypotheses. Whilst the neural network models were high level and developmental in nature, further clarification is necessary in order to transfer findings from an abstract learning system to clinically testable hypotheses. It must be questioned how exactly the neurocomputational parameters relate to biologically plausible mechanisms. For example, can the over-aggressive pruning mechanism map onto biological behaviours? The computational model demonstrated that over-aggressive pruning assesses connections based on their size and strength. However neural pruning in infancy is highly complex in comparison and involves multiple levels of change (Huttenlocher, 2002). In order to compare findings from the computational account, it is essential to clarify the processes that the model is simulating and compare this more rigorously with available biological data.

In Chapter 4, all individual networks were coded by hand based on the trajectory profiles across development, respecting the behavioural grounds in which ASD is diagnosed. Hand coding was chosen as in the initial pruning account (Thomas, Knowland & Karmiloff-Smith, 2011), because of the difficulty in automating the classification of regression or no regression, and into further subgroups against a backdrop of variability in the rates of development in individual networks (what one might call the 'natural variability in

development', here understood to be the consequences of experience-dependent change in parameters in a machine learning system). Whilst individuals demonstrated distinct trajectories when viewed over the development as a whole, a large number of networks displayed oscillations (caused by a high level of internal noise and non-linear underlying mechanisms) particularly at the beginning of development. An aim for future studies would be to employ statistical group modelling techniques to identify subgroups based on curve fitting to developmental trajectories as a complementary subgrouping method. One option would be to use cluster analysis models. As discussed in Chapter 4, one option would to employ a growth trajectory modelling approach, whereby individuals are assumed to fall into a finite amount of groups, each with a distinct trajectory. The model runs with increasing numbers of subgroups until the model fit no longer improves. In such models, the probability of how well an individual fits within a subgroup can also be calculated. This method has been successfully utilised in other developmental disorders such as specific language impairment (Conti-Ramsden, St Clair, Pickles & Durkin, 2012).

6.7 Further study modifications

An important direction for future work stems from the finding in this thesis and prospective studies, that the earliest atypicalities (that have been replicated) are evident at 12 months. Thus far, evidence of atypicalities at 6 months has been inconsistent, with few replicable findings. However, as infants were only tested between at 6 and 12 months, the developmental changes across the first year of infancy are relatively unknown. It would be beneficial to assess infants at smaller intervals in the first 12 months in order to study emerging differences more precisely. The Phase 2 BASIS cohort has tested infants at 5, 10 and 14 months, and I plan, in subsequent work, to utilise these data to examine the first year of infancy for group distinctions, particularly in motor domains.

In the current studies, I did not focus on gender differences between the infants within the BASIS cohort. One prospective study has identified possible sex differences in social engagement tasks at 12 months, arguing for an alternate expression of ASD in females. Chawarksa et al. (2015) reported that female high-risk infants pay more attention to social cues in human faces compared to male at-risk infants at a group level. They proposed that social mechanisms such as this could act as a protective mechanism in females with ASD. However, others have suggested that a similar phenotype in males and females is present in early development (Reinhardt, Wetherby, Schatschneider & Lord, 2015). In future iterations of the predictive statistical model, I would like to investigate potential sex differences, particularly in early development.

In this thesis, statistical logistic and linear regression models were used to identify predictors of outcome. Whilst the models had some success in predicting clinical outcome, more complex modelling techniques could elucidate the relationship between predictors. In Chapter 3, fine motor and AOSI scores were not individually significant predictors of outcome, but did add predictive power to the model as a whole. One possibility is that these factors were involved in higher-order interactions, or were acting as mediating variables. In future iterations of the analyses, I plan to use more sophisticated statistical methods that could better explain the interactions between predictive measures. Unlike linear and logistic regression analyses, techniques such as structural equation modelling (SEM) can analyse all of the predictor variables simultaneously, and therefore disentangle the roles of each predictor. SEM can be considered in hypothesis-driven, confirmatory analyses rather than in exploratory situations, which would be appropriate here. I would also like to run receiver operating characteristic (ROC) curve models to further consider sensitivity and specificity in

the statistical model. The ROC curve is a measure of the model's ability to assign higher probabilities of the outcome to the subgroups that develop the outcome. In other words, we can ask how well the model discriminates between subjects with ASD and without ASD by creating a plot of true positives (sensitivity) and false positives (specificity). Running ROC curve models would allow me to view the changes in sensitivity and specificity as the cut-off probability of the model is changed. This would allow me to consider whether a higher number of true positives or false positives would be more suitable for the theoretical questions surrounding predictive models of ASD.

A number of findings have also provided support for the Over-Pruning account of ASD. Infants exhibited deficits across a range of social and non-social domains. That is, prior to 12 months, development looked seemingly typical for high-risk infants who were later diagnosed with ASD. Social deficits did not precede sensory and motor deficits here. Furthermore, motor deficits identified by coarse measures were apparent when atypicalities were first observed, and these had clinical relevance in predicting later outcome. Higher levels of heterogeneity that were identified at 24 months were associated with poorer social and language outcomes. This finding was not specific to individuals with ASD, but in high-risk infants overall, providing evidence that unaffected infants may display traits common in ASD, but not at the same severity level or across all domains.

6.8 Conclusions

Understanding the interplay between cognitive domains and identifying the earliest behavioural perturbations in infants with ASD can elucidate the patterns of impairments that characterise individuals with the disorder, and help us to understand the underlying causal

mechanisms. In turn, understanding such mechanisms and the changes across development greatly improve the implementation of intervention techniques by identifying targets for treatments, improving screening procedures, and more accurately identifying infants who demonstrate subtle early signs of ASD.

Clinical data from high-risk prospective populations were used to examine the predictive value of intra-subject variability. A more uneven cognitive profile at 24 months was associated with adverse social and language outcomes in some measures at 36 months in high-risk infants. This finding suggests that it may be beneficial to examine variability more specifically in unaffected high-risk infants and consider effects of the broader autism phenotype and the idea of variability as a factor of risk rather than outcome.

In this thesis, I also questioned whether it was possible to predict the diagnostic outcomes of infants based on early behavioural profiles and environmental measures. Atypicalities in multiple measures at 14 months, combined with poorer SES levels could predict diagnostic outcome in a subset of individuals who were later diagnosed with ASD. When this was validated against a second cohort, the results suggest a greater prediction rate of infants who later display more severe symptoms of ASD. It will be of interest to follow these individuals to 36-month outcomes and test the model on a larger sample. This is the first study to utilise a second dataset to validate initial results, which increased the generalisability of my findings. This type of validation is a key challenge for researchers in this field given the heterogeneity present in the disorder. Future studies should aim to identify the extent to which the statistical model can identify a clinical subset of data.

In this thesis, a developmental computational model was used to clarify theoretical discrepancies by simulating development in populations where the underlying causal mechanisms are fully identifiable. Data from the Over-Pruning computational account of ASD was utilised to explore differences in the underlying mechanistic properties of homogenous ASD subgroups, and to assess the implications of intervention, and the underlying variability in outcomes. Four non-regressive subgroups were identified based on developmental trajectories. Here I addressed the unanswered questions that were identified in original the Over-Pruning paper. I was able to identify distinct subgroups in a single population where individual differences (in the form of each subgroup) were identified by a specific profile of neurocomputational parameters, whilst sharing the same underlying pathology in the form of the pruning parameter. I then compared intervention effects in the four non-regressive subgroups using both compensatory and normalisation intervention types. A number of key findings were in line with emerging clinical intervention results. Intervention was more successful when implemented earlier rather than later. There was also evidence of long term, positive intervention effects. Higher pre-intervention ability levels comparable to IQ were predictive of more successful intervention effects. Individual differences demonstrated in the model support the idea of personalised interventions.

Furthermore, I validated hypotheses from the computational model using prospective longitudinal data and my own findings. In order to maximise the impact of intervention strategies, it is vital to understand the developmental processes underlying individual differences in individuals with ASD, and where possible identify more homogenous subgroups. These findings have established testable predictions that can be directly applied to clinical research. This is the first model to advance the developmental theory of

intervention and produce hypotheses about the underlying mechanistic properties in ASD using computational models.

With that in mind, the next stage is to clearly define the abstract learning parameters and identify the most appropriate biological mechanisms that pruning may refer to in the computational data. In terms of clinical findings studies should utilise more fine-grained approaches to measuring early domains, particularly in sensory motor areas will help to clarify the discrepancies in prospective research. The use of more advanced statistical models could also be used to develop current clinical findings, with the potential to provide a clearer picture of the underlying mechanisms in the statistical predictive model used in Chapter 3. Understanding the specificity of mechanistic causes underlying ASD and the origins of variability will be vital to guide interventions to optimise outcomes for the individual child.

Appendix 1. Levene's Test for Equality of Variances

Measure	Levene statistic (*F*)	Sig.
Mullen Gross Motor Raw Score visit 1	3.403	0.077
Mullen Visual Reception Raw Score visit 1	0.636	0.532
Mullen Fine Motor Raw Score visit 1	0.291	0.748
Mullen Receptive Language Raw Score visit 1	3.463	0.089
Mullen Expressive Language Raw Score visit 1	1.204	0.305
Mullen Gross Motor Raw Score visit 2	0.353	0.703
Mullen Visual Reception Raw Score visit 2	0.283	0.754
Mullen Fine Motor Raw Score visit 2	6.49	0.942
Mullen Receptive Language Raw Score visit 2	0.124	0.884
Mullen Expressive Language Raw Score visit 2	2.265	0.109
Mullen Gross Motor Raw Score visit 3	1.553	0.219
Mullen Visual Reception Raw Score visit 3	1.185	0.311
Mullen Fine Motor Raw Score visit 3	0.842	0.434
Mullen Expressive Language Raw Score visit 3	1.044	0.356
Mullen Receptive Language Raw Score visit 3	5.369	0.061
Mullen Visual Reception Raw Score visit 4	**8.377**	**<.005**
Mullen Fine Motor Raw Score visit 4	1.327	0.27
Mullen Receptive Language Raw Score visit 4	**4.729**	**0.011**
Mullen Expressive Language Raw Score visit 4	**2.545**	**0.044**
VABS Communication Domain Standard Score visit 1	3.07	0.069
VABS Daily Living Domain Standard Score visit 1	0.866	0.424
VABS Socialisation Domain Standard Score visit 1	0.041	0.961
VABS Motor Domain Standard Score visit 1	0.056	0.945
VABS Communication Domain Standard Score visit 2	8.654	0.124
VABS Daily Living Domain Standard Score visit 2	1.256	0.29
VABS Socialisation Domain Standard Score visit 2	2.339	0.102

VABS Motor Domain Standard Score visit 2	3.071	0.074
VABS Communication Domain Standard Score visit 3	0.272	0.762
VABS Daily Living Domain Standard Score visit 3	0.549	0.58
VABS Socialisation Domain Standard Score visit 3	1.886	0.157
VABS Motor Domain Standard Score visit 3	0.838	0.436
VABS Communication Domain Standard Score visit 4	3.199	0.074
VABS Daily Living Domain Standard Score visit 4	2.319	0.104
VABS Socialisation Domain Standard Score visit 4	4.022	0.121
VABS Motor Domain Standard Score visit 4	2.562	0.096

Table A.1. Significant and Non-significant and Levene's test statistics for Mullen and Vineland behavioural measures at 7,14,24 and 36 months.

Appendix 2. Parental demographic information for Phase 1 and Phase 2

	HR Siblings		
	Combined	No ASD	ASD
Phase 1			
Household Income			
1	4	4	0
2	15	10	5
3	19	10	9
4	6	3	3
5	8	8	0
Father's Qualification			
1	10	7	3
2	10	7	3
3	24	15	9
4	6	5	1
5	0	0	0
Father's Job			
1	3	2	1
2	10	10	0
3	22	11	11
4	11	6	5
5	5	5	0

Mother's Qualification			
1	5	3	2
2	18	10	8
3	15	12	5
4	10	8	2
5	1	1	0
Mother's Job			
1	27	16	11
2	7	5	2
3	8	5	3
4	5	3	2
5	1	1	0

	HR Siblings		
	Combined	No-ASD	ASD
Phase 2			
Household Income			
1	10	7	3
2	31	27	4
3	25	19	6
4	11	11	0
5	21	20	1
Father's Qualification			
1	12	11	1
2	31	27	4
3	30	24	6
4	16	1	15
5	6	4	2
Father's Job			
1	8	6	2
2	30	24	6
3	41	39	2
4	20	16	4
5	1	1	0
Mother's Qualification			
1	13	10	3
2	23	17	6
3	40	37	3
4	18	17	1
5	6	6	0

Mother's Job			
1	39	31	8
2	22	20	2
3	23	20	2
4	17	16	1
5	0	0	0

A.2 .Demographic information for Phase 1 and 2 low-risk and high-risk groups.

Appendix 3. Probability distributions for remaining parameters in HR, HREO and HREOC populations.

Hidden Units	**10**	**20**	**30**	**40**	**50**	**60**	**75**	**100**	**200**	**350**
High-risk (%)	0.1	0.8	3	13.5	21.2	25.6	19.7	10.7	4	1
High-risk early onset (%)	0	0.7	4.1	11.3	21	22.8	22.5	13.6	4.2	0.2
High-risk early onset control (%)	0	0.7	4.1	11.3	21	22.8	22.5	13.6	4.2	0.2

Table A.3.1. Probability distributions for Hidden Units parameters for individuals in HR, HREO and HREOC populations.

Temperature	**0.0625**	**0.125**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**2**	**3**	**4**
High-risk (%)	0.3	1	4.6	13.3	19.6	23.7	18.6	12.9	3.8	0.9	0.1
High-risk early onset (%)	0	0.2	4	14.3	21.5	22.8	17.8	12.2	5.5	1	0
High-risk early onset control (%)	0	0.2	4	14.3	21.5	22.8	17.8	12.2	5.5	1	0

Table A.3.2 Probability distributions for Temperature parameters for individuals in HR, HREO and HREOC populations.

Noise	**0**	**0.05**	**0.1**	**0.2**	**0.5**	**2**	**4**	**6**

High-risk (%)	2.7	10	22	30.9	19.4	11.9	2.5	0.2
High-risk early onset (%)	4.1	10.5	23.6	26.7	21.4	9.3	3.3	0.3
High-risk early onset control (%)	4.1	10.5	23.6	26.7	21.4	9.3	3.3	0.3

Table A.3. Probability distributions for Noise parameters for individuals in HR, HREO and HREOC populations.

Learning Rate	**0.005**	**0.01**	**0.025**	**0.05**	**0.075**	**0.1**	**0.125**	**0.15**	**0.175**	**0.25**	**0.3**
High-risk (%)	0	0.1	1.9	5.8	9.6	20.4	24.3	18	12.7	0	0
High-risk early onset (%)	0	0	1.8	5.3	11.9	15.6	23.7	23.3	10.5	1.3	0.2
High-risk early onset control (%)	0	0	1.8	5.3	11.9	15.6	23.7	23.3	10.5	1.3	0.2

Table A.3.3 Probability distributions for Learning Rate parameters for individuals in HR, HREO and HREOC populations.

Momentum	**0**	**0.05**	**0.1**	**0.15**	**0.2**	**0.35**	**0.5**	**0.6**	**0.75**
High-risk (%)	0.4	3.7	11.6	20.8	27.2	20.3	9.7	4.4	0.4
High-risk early onset (%)	0.2	3.9	10.3	21.1	27.3	20.3	12.6	3.7	0.1
High-risk early onset control (%)	0.2	3.9	10.3	21.1	27.3	20.3	12.6	3.7	0.1

Table A.3.4 Probability distributions for Momentum parameters for individuals in HR, HREO and HREOC populations.

Weight variance	**0.01**	**0.05**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**
High-risk (%)	0.3	2.8	12	18.2	26.1	25.1	12.6	22	0.2
High-risk early onset (%)	0.3	2.1	8.9	20.7	29.4	23.9	9.6	4.2	0.4
High-risk early onset control (%)	0.3	2.1	8.9	20.7	29.4	23.9	9.6	4.2	0.4

Table A.3.5 Probability distributions for Weight Variance parameters for individuals in HR, HREO and HREOC populations.

Architecture	**0**	**1**	**2**
High-risk (%)	9.3	78.1	12.5
High-risk early onset (%)	79.4	10.8	9.6
High-risk early onset control (%)	79.4	10.8	9.6

Table A.3.6 Probability distributions for Architecture parameters for individuals in HR, HREO and HREOC populations.

Learning Algorithm	**0**	**1**
High-risk (%)	5.9	94
High-risk early onset (%)	6.6	93.4
High-risk early onset control (%)	6.6	93.4

Table A.3.7 Probability distributions for learning algorithm parameters for individuals in the HR, HREO and HREOC populations.

NN Threshold	**0.0025**	**0.005**	**0.01**	**0.025**	**0.1**	**0.15**	**0.2**	**0.25**	**0.5**
High-risk (%)	0.02	1.1	3.5	12.2	43.8	20.8	11.7	4.5	1.8
High-risk early onset (%)	0.02	0.06	4.4	13.7	45.3	20.5	9.2	4	1.6
High-risk early onset control (%)	0.02	0.06	4.4	13.7	45.3	20.5	9.2	4	1.6

Table A.3.8 Probability distributions for NN Threshold parameters for individuals in HR, HREO and HREOC populations.

Pruning Probability	**0**	**0.01**	**0.025**	**0.05**	**0.1**	**0.5**	**0.75**	**1**

High-risk (%)	0.2	2.6	10	48.3	23.2	10	3.2	0.4
High-risk early onset (%)	0.6	3.3	12.4	46.3	23.5	10.1	2.6	0.7
High-risk early onset control (%)	0.6	3.3	12.4	46.3	23.5	10.1	2.6	0.7

Table A.3.9 Probability distributions for Pruning Probability parameters for individuals in HR, HREO and HREOC populations.

Sparseness	**0**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	0.5
High-risk (%)	60.4	20.1	11.21	5.2	2.3	0.3	0.2
High-risk early onset (%)	62.7	18.2	12.3	5.2	1.2	0.2	0
High-risk early onset control (%)	62.7	18.2	12.3	5.2	1.2	0.2	0

Table A.3.10 Probability distributions for Sparseness parameters for individuals in HR, HREO and HREOC populations.

References

Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews Genetics*, *9*(5), 341-355.

Adrien, J. L., Lenoir, P., Martineau, J., Perrot, A., Hameury, L., Larmande, C., & Sauvage, D. (1993). Blind ratings of early symptoms of autism based upon family home movies. *Journal of the American Academy of Child & Adolescent Psychiatry*, *32*(3), 617-626

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). Washington, DC

Anderson, D. K., Liang, J. W., & Lord, C. (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *55*(5), 485-494.

Anderson, D. K., Lord, C., Risi, S., Shulman, C., Welch, K., DiLavore, P. S., Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology, 75*(4), 594–604.

Andrew, P., & Gustafsson, L. (2003). Detailed learning in narrow fields–towards a neural network model of autism. In *Artificial Neural Networks and Neural Information Processing* (pp. 830–838). Springer.

Auyeung, Bonnie, et al. "Fetal testosterone and autistic traits." *British Journal o Psychology 100.1* (2009): 1-22.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S. (2000). A screening instrument for autism at 18 month of age: a six-year followup study. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 694-702.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., et al. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *Lancet, 368,* 210–215.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of developmental psychology*, *4*(2), 113-125.

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., & Johnson, M. H. (2012). Precursors to social and communication difficulties in infants at-risk for autism: Gaze following and attentional engagement. *Journal of autism and developmental disorders*, *42*(10), 2208-2218.

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in cognitive sciences*, *10*(6), 258-264.

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *The Journal of Neuroscience*, *24*(42), 9228-9231.

Belmonte, M. K., & Yurgelun-Todd, D. A. (2003). Functional anatomy of impaired selective attention and compensatory processing in autism. *Cognitive brain research*, *17*(3), 651-664.

Beversdorf, D. Q., Narayanan, A., Hillier, A., & Hughes, J. D. (2007). Network model of decreased context utilization in autism spectrum disorder. *Journal of autism and developmental disorders*, *37*(6), 1040-1048.

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *The Journal of Neuroscience*, *24*(42), 9228-9231.

Bhat, A. N., Galloway, J. C., & Landa, R. J. (2012). Relation between early motor delay and later communication delay in infants at risk for autism. *Infant Behavior and Development*, *35*(4), 838-846.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., et al. (1994). A case-control family study of autism. *Journal of Child Psychology and Psychiatry, 35,* 877–900.

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., ... & Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, *48*(9), 2465-2475.

Boucher, J. (2011). Redefining the Concept of Autism as a Unitary Disorder: Multiple Causal Deficits of a Single Kind?. *The neuropsychology of autism*, 469.

Brandler, W. M., & Sebat, J. (2015). From De Novo Mutations to Personalized Therapeutic Interventions in Autism. *Annual review of medicine*, *66*, 487-507.

Bradley, R. H., & Corwyn, R. F. (2004). Life satisfaction among European American, African American, Chinese American, Mexican American, and Dominican American adolescents. *International Journal of Behavioral Development*, *28*(5), 385-400.

Bradshaw, J., Steiner, A. M., Gengoux, G., & Koegel, L. K. (2015). Feasibility and effectiveness of very early intervention for infants at-risk for autism spectrum disorder: A systematic review. *Journal of autism and developmental disorders*, *45*(3), 778-794.

Braitenberg, V. (2001). Brain size and number of neurons: An exercise in synthetic neuroanatomy. *Journal of Computational Neuroscience*, *10(1*), 71–77.

Brock, J., Brown, C. C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Development and psychopathology*, *14*(02), 209-224.

Brugha, T et al. "Estimating the prevalence of autism spectrum conditions in adults: Extending the 2007 adult psychiatric morbidity survey." *The Health and Social Care Information Centre (NHS)* (2012).

Bryson, S. E., Zwaigenbaum, L., Brian, J., Roberts, W., Szatmari, P., Rombough, V., & McDermott, C. (2007). A prospective case series of high-risk infants who developed autism. *Journal of autism and developmental disorders*, *37*(1), 12-24.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The Autism Observation Scale for Infants: scale development and reliability data. *Journal of autism and developmental disorders*, *38*(4), 731-738.

Burns, T. G., King, T. Z., & Spencer, K. S. (2013). Mullen scales of early learning: the utility in assessing children diagnosed with autism spectrum disorders, cerebral palsy, and epilepsy. *Applied Neuropsychology: Child*, *2*(1), 33-42.

Caron, M. J., Mottron, L., Berthiaume, C., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinnings of visuospatial peaks in autism. *Brain*, *129*(7), 1789-1802.

Carr, E. G., & Dores, P. A. (1981). Patterns of language acquisition following simultaneous communication with autistic children. *Analysis and Intervention in Developmental Disabilities*, *1*(3), 347-361.

Carr, T., Shih, W., Lawton, K., Lord, C., King, B., & Kasari, C. (2015). The relationship between treatment attendance, adherence, and outcome in a caregiver-mediated intervention for low-resourced families of young children with autism spectrum disorder. *Autism*,

Casanova, M. F., van Kooten, I. A., Switala, A. E., van Engeland, H., Heinsen, H., Steinbusch, H. W., ... & Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta neuropathologica*, *112*(3), 287-303.

Cassel, T. D., Messinger, D. S., Ibanez, L. V., Haltigan, J. D., Acosta, S. I., & Buchman, A. C. (2007). Early social and emotional communication in the infant siblings of children with autism spectrum disorders: An examination of the broad phenotype. *Journal of autism and developmental disorders*, *37*(1), 122-132.

Charman, T. (2014). Variability in Neurodevelopmental Disorders. *Neurodevelopmental Disorders: Research Challenges and Solutions*, 117.

Charman, T. and Baird, G. (2002). Practitioner review: Assessment and diagnosis of autism spectrum disorders in the pre-school years. *Journal of Child Psychology and Psychiatry, 43,* 289–305.

Charman, T., Jones, C. R., Pickles, A., Simonoff, E., Baird, G., & Happé, F. (2011). Defining the cognitive phenotype of autism. *Brain research*, *1380*, 10-21.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J. A., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry*, *46*(5), 500-513.

Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry*, *50*(10), 1235-1245.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry*, *48*(2), 128-138.

Chawarska, K., Macari, S., Powell, K., DiNicola, L., & Shic, F. (2015). Enhanced Social Attention in Female Infant Siblings at Risk for Autism. *Journal of the American Academy of Child & Adolescent Psychiatry*.

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological psychiatry*, *74*(3), 195-203.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231-239.

Chonnaparamutt, W., & Barakova, E. I. (2008). Robot Simulation of Sensory Integration Dysfunction in Autism with Dynamic Neural Fields Model. In *Artificial Intelligence and Soft Computing* (pp. 741–751). Springer.

Church, B. A., Krauss, M. S., Lopata, C., Toomey, J. A., Thomeer, M. L., Coutinho, M. V., Volker, M. A., & Mer- cado, E. (2010). Atypical categorization in children with high-functioning autism spectrum disorder. *Psychonomic Bulletin & Review*, *17*, 862–868.

Clifford, S. M., Hudry, K., Elsabbagh, M., Charman, T., Johnson, M. H., & BASIS Team. (2013). Temperament in the first 2 years of life in infants at high-risk for autism spectrum disorders. *Journal of autism and developmental disorders*, *43*(3), 673-686.

Cohen, I.L. (1994). An artificial neural network analogue of learning in autism. Biological Psychiatry, 36, 5–20.

Cohen, I.L. (1998). Neural network analysis of learning in autism. In D.J. Stein & J. Ludik (Eds.), Neural networks and psychopathology (pp. 274 –315). New York: Cambridge University Press

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, *12*(4), 425-434.

Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N. and Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 719–726.

Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M. and Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry, 167,* 1349–1356.

Conti-Ramsden, G., St Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental rajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1716-1735.

Cook Jr, E. H., & Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature*, *455*(7215), 919-923.

Courchesne, E, et al. "Unusual brain growth patterns in early life in patients with autistic disorder an MRI study." *Neurology* 57.2 (2001): 245-254.

Courchesne, E., & Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current opinion in neurobiology*, *15*(2), 225-230.

Courchesne, E., & Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current*

opinion in neurobiology, *15*(2), 225-230.

Damiano, C. R., Aloi, J., Treadway, M., Bodfish, J. W., & Dichter, G. S. (2012). Adults with autism spectrum disorders exhibit decreased sensitivity to reward parameters when making effort-based decisions. *Journal of neurodevelopmental disorders*, *4*(1), 13.

Dawson, G., Jones, E. J., Merkle, K., Venema, K., Lowy, R., Faja, S., ... & Smith, M. (2012). Early behavioral intervention is associated with normalized brain activity in young children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(11), 1150-1159.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of autism and developmental disorders*, *28*(6), 479-485.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., ... & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*, *125*(1), e17-e23.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental neuropsychology*, *27*(3), 403-424.

De Carvalho, L. A. V., De Carvalho Ferreira, N., & Fiszman, A. (1999). A

neurocomputational model for autism. In *Proceedings of IV Brazilian conference on neural networks-IV* (pp. 888–999).

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., ... & Singh, T. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209-215.

DiLalla, D. L., & Rogers, S. J. (1994). Domains of the Childhood Autism Rating Scale: Relevance for diagnosis and treatment. *Journal of Autism and Developmental Disorders*, *24*(2), 115-128.

Di Martino, A., Zuo, X. N., Kelly, C., Grzadzinski, R., Mennes, M., Schvarcz, A., ... & Milham, M. P. (2013). Shared and distinct intrinsic functional network centrality in autism and attention-deficit/hyperactivity disorder. *Biological psychiatry*, *74*(8), 623-632.

Dovgopoly, A., & Mercado III, E. (2013). A connectionist model of category learning by individuals with high- functioning autism spectrum disorder. *Cognitive, Affective, & Behavioral Neuroscience*, *13*, 371–389.

Duch, W., Dobosz, K., & Mikołajewski, D. (2013). Autism and ADHD–two ends of the same spectrum? In *Neural Information Processing* (pp. 623–630). Springer.

Duch, W., Nowak, W., Meller, J., Osin´ski, G., Dobosz, K., Mikołajewski, D., & Wo´jcik, G. M. (2012). Computational approach to understanding autism

spectrum disorders. *Computer Science*, *13*, 47–61.

Ebbels, S. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. Child Language Teaching and Therapy, 30(1), 7-40.

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., Johnson, M. H., & British Autism Study of Infant Siblings Team. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*(3), 189-194.

Elsabbagh, M., Gliga, T., Pickles, A., Hudry, K., Charman, T., Johnson, M. H., & BASIS Team. (2013). The development of face orienting mechanisms in infants at-risk for autism. *Behavioural Brain Research*, *251*, 147-154.

Elsabbagh, M., Volein, A., Csibra, G., Holmboe, K., Garwood, H., Tucker, L., ... & Baird, G. (2009). Neural correlates of eye gaze processing in the infant broader autism phenotype. *Biological psychiatry*, *65*(1), 31-38.

Elsabbagh, M. and Johnson, M. H. (2010). Getting answers from babies about autism. *Trends in Cognitive Sciences, 14,* 81-87.

Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., ... & Gerig, G. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *American Journal of Psychiatry*.

Estes, A. M., Dawson, G., Sterling, L., & Munson, J. (2007). Level of intellectual functioning predicts patterns of associated symptoms in school-age children with autism spectrum disorder. *American Journal on Mental Retardation*, *112*(6), 439-449.

Focaroli, V., Taffoni, F., Parsons, S. M., Keller, F., & Iverson, J. M. (2016). Performance of motor sequences in children at heightened vs. low risk for ASD: A longitudinal study from 18 to 36 months of age. *Frontiers in psychology*, *7*.

Fein, D., Barton, M., Eigsti, I. M., Kelley, E., Naigles, L., Schultz, R. T., ... & Troyb, E. (2013). Optimal outcome in individuals with a history of autism.*Journal of Child Psychology and Psychiatry*, *54*(2), 195-205.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192-195.

Flanagan, J. E., Landa, R., Bhat, A., & Bauman, M. (2012). Head lag in infants at risk for autism: a preliminary study. *American Journal of Occupational Therapy*, *66*(5), 577-585.

Frith, C. (2003). What do imaging studies tell us about the neural basis of autism. *Autism: Neural basis and treatment possibilities*, 149-176.

Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Blackwell.

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind". *Cognition*,*50*(1), 115-132.

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., ... & Ripke, S. (2014). Most genetic risk for autism resides with common variation. *Nature genetics*, *46*(8), 881-885.

Georgiades, S., Szatmari, P., Zwaigenbaum, L., Bryson, S., Brian, J., Roberts, W., ... & Garon, N. (2013). A prospective study of autistic-like traits in unaffected siblings of probands with autism spectrum disorder. *JAMA psychiatry*, *70*(1), 42-48.

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. *Current opinion in neurobiology*,*17*(1), 103-111

Geschwind, D. H. State MW (2015) Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurology*, *14*(1109), 00044-7.

Gillberg, C., Ehlers, S., Schaumann, H., Jakobsson, G., Dahlgren, S.O., Lindblom, R., Bagenholm, A., Tjuus, T., Blidner, E., 1990. Autism under age 3 years: a clinical study of 28 cases referred for autistic symptoms in infancy. *Journal of Child Psycholy & Psychiatry 31,* 921–934.

Gliga, T., Bedford, R., Charman, T., Johnson, M. H., & BASIS Team. (2015). Enhanced visual search in infancy predicts emerging autism symptoms. *Current Biology*, *25*(13), 1727-1730.

Gordon, R. G., & Watson, L. R. (2015). Brief Report: Gestures in Children at Risk for Autism Spectrum Disorders. *Journal of autism and developmental disorders*, *45*(7), 2267-2273.

Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., ... & Barrett, B. (2010). Parent-mediated communication-focused treatment in children with autism (PACT): a randomised controlled trial. *The Lancet*, *375*(9732), 2152-2160.

Green, J., Charman, T., Pickles, A., Wan, M. W., Elsabbagh, M., Slonims, V., ... & Jones, E. J. (2015). Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial. *The Lancet Psychiatry*, *2*(2), 133-140.

Grossberg, S., & Seidman, D. (2006). Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates. *Psychological review*, *113*(3), 483.

Gustafsson, L. (1997). Inadequate cortical feature maps: A neural circuit theory of autism. *Biological psychiatry*, *42*(12), 1138-1147.

Gustafsson, L., & Paplin´ski, A. P. (2004). Self-organization of an artificial neural network subjected to attention shift impairments and familiarity preference, characteristics studied in autism. *Journal of Autism and Developmental Disorders*, *34*, 189–198.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in cognitive sciences*, *13*(2), 65-73.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews Neuroscience*, *11*(9), 651-659.

Hadjikhani, N., Joseph, R. M., Snyder, J., & Tager-Flusberg, H. (2006). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral cortex*, *16*(9), 1276-1282.

Hale, J. B., Chen, S. A., Tan, S. C., Poon, K., Fitzer, K. R., & Boyd, L. A. (2016). Reconciling individual differences with collective needs: The juxtaposition of sociopolitical and neuroscience perspectives on remediation and compensation of student skill deficits. *Trends in Neuroscience and Education*, *5*(2), 41-51.

Halit, H., Csibra, G., Volein, A., & Johnson, M. H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry*, *45*(7), 1228-1234.

Hardan, A. Y., Libove, R. A., Keshavan, M. S., Melhem, N. M., & Minshew, N. J. (2009). A preliminary longitudinal magnetic resonance imaging study of brain volume and cortical thickness in autism. *Biological psychiatry*, *66*(4), 320-326.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, *36*(1), 5-25.

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research.*Neuropsychology review*, *18*(4), 287-304.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature neuroscience*, *9*(10), 1218-1220.

Hedvall, Å., Fernell, E., Holm, A., Åsberg Johnels, J., Gillberg, C., & Billstedt, E. (2013). Autism, processing speed, and adaptive functioning in preschool children. *The Scientific World Journal*, *2013*.

Hoshino, Y., Kaneko, M., Yashima, Y., Kumashiro, H., Volkmar, F. R., & Cohen, D. J. (1987). Clinical features of autistic children with setback course in their infancy. *Psychiatry and Clinical Neurosciences*, *41*(2), 237-245.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*, *45*(2), 212-229.

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American journal on intellectual and developmental disabilities*, *114*(1), 23-41.

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., John- son, M., Charman, T., the BASIS team, 2013. Early language profiles in infants at high-risk for autism spectrum disorders. J. Autism Dev. Disord.,

Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H. Aldred, C., et al. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International Journal of Language and Communication Disorders, 45,* 681-690.

Huttenlocher, P.R. (2002). Neural plasticity: The effects of environment on the development of the cerebral cortex. Cambridge, MA: Harvard University Press.

Huttenlocher, P. R. (2009). *Neural plasticity*. Harvard University Press.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis In human cerebral cortex. *Journal of comparative Neurology*,*387*(2), 167-178.

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of autism and developmental disorders*, *37*(1), 158-170

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*(7), 475-483.

Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development?. *Developmental Cognitive Neuroscience*, *1*(1), 7-21.

Johnson, M. H., Grossmann, T., & Kadosh, K. C. (2009). Mapping functional brain development: Building a social brain through interactive specialization.*Developmental psychology*, *45*(1), 151.

Jones, E. J., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: a review of prospective studies of infants at risk. *Neuroscience & Biobehavioral Reviews*, *39*, 1-33.

Jones, C. R., Happé, F., Baird, G., Simonoff, E., Marsden, A. J., Tregay, J., ... & Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(13), 2850-2858.

Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: evidence from an FMRI study of an executive function task and corpus callosum morphometry. *Cerebral cortex*, *17*(4), 951-961.

Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K., & Varma, S. (2012). Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1292-1313.

Kaale, A., Smith, L., & Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism. *Journal of Child Psychology and Psychiatry*, *53*(1), 97-105.

Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2006). Sentence comprehension in autism: thinking in pictures with decreased functional connectivity. *Brain*, *129*(9), 2484-2493.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in cognitive sciences*, *2*(10), 389-398.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental psychology*, *45*(1), 56.

Kasari, Connie, Stephanny FN Freeman, and Tanya Paparella. "Early intervention in autism: Joint attention and symbolic play." *International review of research in mental retardation* 23 (2000): 207-237.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry*, *47*(6), 611-620.

Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of autism and developmental disorders*, *40*(9), 1045-1056.

Kasari, C., Lawton, K., Shih, W., Barker, T. V., Landa, R., Lord, C., ... & Senturk, D. (2014). Caregiver-mediated intervention for low-resourced preschoolers with autism: an RCT. *Pediatrics*, *134*(1), e72-e79.

Kawakubo, Y., Kasai, K., Okazaki, S., Hosokawa-Kakurai, M., Watanabe, K. et al. (2007). Electrophysiological abnor- malities of spatial attention in adults with autism during the gap overlap task. Clinical Neurophysiology: Official Journal of the International Federation of Clinical *Neurophysiology, 118,* 1464–1471.

Keary, C. J., Minshew, N. J., Bansal, R., Goradia, D., Fedorov, S., Keshavan, M. S., & Hardan, A. Y. (2009). Corpus callosum volume and neurocognition in autism. *Journal of autism and developmental disorders*, *39*(6), 834-841.

Kim, S. H., Macari, S., Koller, J., & Chawarska, K. (2016). Examining the phenotypic heterogeneity of early autism spectrum disorder: subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry*, *57*(1), 93-102.

Klin, A., Chawarska, K., Paul, R., Rubin, E., Morgan, T., Wiesner, L., & Volkmar, F. (2004). Autism in a 15-month-old child. *American Journal of Psychiatry*, *161*(11), 1981-1988.

Koshino, H., Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2008). fMRI investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. *Cerebral cortex*, *18*(2), 289-300.

Kriete, T., & Noelle, D. C. (2015). Dopamine and the development of executive dysfunction in autism spectrum disorders. *PLoS ONE*, *10(3)*, e0121605.

Lambert-Brown, B. L., McDonald, N. M., Mattson, W. I., Martin, K. B., Ibañez, L. V., Stone, W. L., & Messinger, D. S. (2015). Positive emotional engagement and autism risk. *Developmental psychology*, *51*(6), 848.

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry*, *47*(6), 629-638.

Landa, R. J., Gross, A. L., Stuart, E. A., & Bauman, M. (2012). Latent class analysis of early developmental trajectory in baby siblings of children with autism. *Journal of Child Psychology and Psychiatry*, *53*(9), 986-996.

Landa, R. J., Gross, A. L., Stuart, E. A., & Faherty, A. (2013). Developmental trajectories in children with and without autism spectrum disorders: the first 3 years. *Child development*, *84*(2), 429-442.

Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of general psychiatry*, *64*(7), 853-864.

Landry, R., & Bryson, S. E. (2004). Impaired disengagement of attention in young children with autism. *Journal of Child Psychology and Psychiatry*, *45*(6),

1115-1122.

Langen, M., Schnack, H. G., Nederveen, H., Bos, D., Lahuis, B. E., de Jonge, M. V., & Durston, S. (2009). Changes in the developmental trajectories of striatum in autism. *Biological psychiatry*, *66*(4), 327-333.

Larsson, M., Weiss, B., Janson, S., Sundell, J., & Bornehag, C. G. (2009). Associations between indoor environmental factors and parental-reported autistic spectrum disorders in children 6–8 years of age. *Neurotoxicology*, *30*(5), 822-831.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Front Hum Neurosci*, *8*.

Lazenby, D. C., Sideridis, G. D., Huntington, N., Prante, M., Dale, P. S., Curtin, S., ... & Akshoomoff, N. (2015). Language Differences at 12 Months in Infants Who Develop Autism Spectrum Disorder. *Journal of autism and developmental disorders*, 1-11.

Leonard, H. C., Bedford, R., Charman, T., Elsabbagh, M., Johnson, M. H., Hill, E. L., & Holmboe, K. (2014). Motor development in children at risk of autism: a follow-up study of infant siblings. *Autism*, *18*(3), 281-291.

Levy, H. M. (2008). Meeting the needs of all students through differentiated instruction: Helping every child reach and exceed standards. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *81*(4), 161-164.

Lewis, J. D., & Elman, J. L. (2008). Growth-related neural reorganization and the autism phenotype: a test of the hypothesis that altered brain growth leads to altered connectivity. *Developmental science*, *11*(1), 135-155.

Lewis, M. H., Tanimura, Y., Lee, L. W., & Bodfish, J. W. (2007). Animal models of restricted repetitive behavior in autism. *Behavioural brain research*, *176*(1), 66-74.

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of general psychiatry*, *63*(6), 694-701.

Lord, C., Risi, S., Lambrecht, L., Cook Jr, E. H., Leventhal, B. L., DiLavore, P. C., ... & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, *30*(3), 205-223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, *24*(5), 659-685.

Lord, C., Shulman, C., & DiLavore, P. (2004). Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*,*45*(5), 936-955.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of autism and developmental disorders*, *38*(8), 1426-1438.

Luyster, R., Richler, J., Risi, S., Hsu, W. L., Dawson, G., Bernier, R., ... & Goudie-Nice, J. (2005). Early regression in social communication in autism spectrum disorders: a CPEA Study. *Developmental neuropsychology*, *27*(3), 311-336.

Mareschal, D., & Thomas, M. S. (2007). Computational modeling in developmental psychology. *Evolutionary Computation, IEEE Transactions on*, *11*(2), 137-150

Marcus, L. M., Lansing, M., & Schopler, E. (1993). Assessment of children with autism and pervasive developmental disorder. In J. L. Culbertson & D. J. Willis (Eds.), *Testing young children: A reference guide for developmental, psychoeducational, and psychosocial assessments* (pp. 319–3

Mason, R. A., Williams, D. L., Kana, R. K., Minshew, N., & Just, M. A. (2008). Theory of mind disruption and recruitment of the right hemisphere during narrative comprehension in autism. *Neuropsychologia*, *46*(1), 269-280.

Mundy, P., & Neal, R. (2001). Neural plasticity, joint attention and autistic developmental pathology. *International review of research in mental retardation*, *23*, 139-168.

McClelland, J. L. (2000). The basis of hyperspecificity in autism: A preliminary suggestion based on properties of neural nets. *Journal of autism and developmental disorders*, *30*(5), 497-502.

Milne, E., White, S., Campbell, R., Swettenham, J., Hansen, P., & Ramus, F. (2006). Motion and form coherence detection in autistic spectrum disorder: Relationship to motor control and 2: 4 digit ratio. *Journal of autism and developmental disorders*, *36*(2), 225-237.

Milne, M., Luerssen, M., Lewis, T., Leibbrandt, R., & Powers, D. (2011). Designing and evaluating interactive agents as social skills tutors for children with autism spectrum disorder. *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*, 23-48.

Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., Smith, I., & Bryson, S. (2006). Early language and communication development of infants later diagnosed with autism spectrum disorder. *Journal of Developmental & Behavioral Pediatrics*, *27*(2), S69-S78.

Monk, C. S., Peltier, S. J., Wiggins, J. L., Weng, S. J., Carrasco, M., Risi, S., & Lord, C. (2009). Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage*, *47*(2), 764-772.

Mottron, L., and Burack, J.A. (2001). *Enhanced perceptual functioning in the development of autism.* In J.A. Burack, T. Charman, N. Yirmiya, and P.R. Zelazo (Eds.), The development of autism: Perspectives from theory and research (pp. 131–148). Mahwah, NJ: Lawrence Erlbaum Associates.

Mottron, L., Burack, J. A., Iarocci, G., Belleville, S. and Enns, J. T. (2003). Locally oriented perception with intact global processing among adolescents with high-functioning autism: evidence from multiple paradigms. *Journal of Child Psychology and Psychiatry, 44,* 904–13.

Mullen, E. M. (1995). *Mullen scales of early learning* (pp. 58-64). Circle Pines, MN: AGS.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (Vol. 1). Psychology Press.

Mundy, P., & Neal, R. (2001). Neural plasticity, joint attention and autistic developmental pathology. *International review of research in mental retardation*, *23*, 139-168.

Nele, D., Ellen, D., Petra, W., & Herbert, R. (2015). Social information processing in infants at risk for ASD at 5 months of age: The influence of a familiar face and direct gaze on attention allocation. *Research in Autism Spectrum Disorders*, *17*, 95-105.

Nomi, J. S., & Uddin, L. Q. (2015). Developmental changes in large-scale network connectivity in autism. *NeuroImage: Clinical*, *7*, 732-741.

Noriega, G. (2008b). A neural model for compensation of sensory abnormalities in autism through feedback from a measure of global perception. *Neural Networks, IEEE Transactions on*, *19*, 1402–1414.

Ohta, M., Nagai, Y., Hara, H., Sasaki, M., 1987. Parental perception of behavioral symptoms in Japanese autistic children. J. Autism Dev. Disord. 17, 549–563.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of autism and developmental disorders*, *24*(3), 247-257.

Ozonoff, S., Iosif, A. M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., ... & Steinfeld, M. B. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(3), 256-266.

Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: patterns of symptom emergence in the first years of life.*Autism research*, *1*(6), 320-328.

Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: patterns of symptom emergence in the first years of life. *Autism research*, *1*(6), 320-328.

Ozonoff, S., Heung, K., & Thompson, M. (2011). Regression and other patterns of onset. *Autism spectrum disorders*, 60-74.

Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., ... & Steinfeld, M. (2014). The broader autism phenotype in infancy: when does it emerge?. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(4), 398-407.

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., ... & Hutman, T. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*, *128*(3), e488-e495.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... & Zwaigenbaum, L. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, *56*(9), 988-998.

Pardo, C. A., & Eberhart, C. G. (2007). The neurobiology of autism. *Brain Pathology*, *17*(4), 434-447.

Patterson, K., Plaut, D. C., Mcclelland, J. L., Seidenberg, M. S., Behrmann, M., & Hodges, J. R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. *Neural modeling of brain and cognitive disorders*, 177-199.

Pearson, R. K. (2002). Outliers in process modeling and identification. *Control Systems Technology, IEEE Transactions on*, *10*(1), 55-63.

Peterson, D., Mahajan, R., Crocetti, D., Mejia, A., & Mostofsky, S. (2015). Left-Hemispheric Microstructural Abnormalities in Children With High-Functioning Autism Spectrum Disorder. *Autism Research*, *8*(1), 61-72.

Persico, A. M., & Bourgeron, T. (2006). Searching for ways out of the autism maze: Genetic, epigenetic and environmental clues. Trends in Neurosciences, 29(7), 349–358.

Pickles, A., Starr, E., Kazak, S., Bolton, P., Papanikolaou, K., Bailey, A., ... & Rutter, M. (2000). Variable expression of the autism broader phenotype: findings from extended pedigrees. *Journal of Child Psychology and Psychiatry*, *41*(04), 491-502.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry*, *154*(2), 185-190.

Presmanes, A. G., Walden, T. A., Stone, W. L., & Yoder, P. J. (2007). Effects of different attentional cues on responding to joint attention in younger siblings of children with autism spectrum disorders. *Journal of autism and developmental disorders*, *37*(1), 133-144.

Ratajczak, H. V. (2011). Theoretical aspects of autism: Causes-A review.*Journal of immunotoxicology*, *8*(1), 68-79.

Reinhardt, V. P., Wetherby, A. M., Schatschneider, C., & Lord, C. (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. *Journal of autism and developmental disorders*, *45*(3), 697-706.

Revithis, S., & Tagalakis, G. (2012). A SOM-Based validation approach to a neural circuit theory of autism. In *Artificial Intelligence: Theories and Applications* (pp. 25–32). Springer.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Rogers, S. J., & DiLalla, D. L. (1990). Age of symptom onset in young children with pervasive developmental disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *29*(6), 863-872.

Rogers, S. J., Vismara, L., Wagner, A. L., McCormick, C., Young, G., & Ozonoff, S. (2014). Autism treatment in the first year of life: a pilot study of infant start, a parent-implemented intervention for symptomatic infants. *Journal of autism and developmental disorders*, *44*(12), 2981-2995.

Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. *Developmental science*, *8*(5), 444-458.

Ronald, A., Happé, F., Price, T. S., Baron-Cohen, S., & Plomin, R. (2006). Phenotypic and genetic overlap between autistic traits at the extremes of the general population. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(10), 1206-1214.

Rowberry, J., Macari, S., Chen, G., Campbell, D., Leventhal, J. M., Weitzman, C., & Chawarska, K. (2015). Screening for autism spectrum disorders in 12-month-old high-risk siblings by parental report. *Journal of autism and developmental disorders*, *45*(1), 221-229.

Rozga, A., Hutman, T., Young, G. S., Rogers, S. J., Ozonoff, S., Dapretto, M., & Sigman, M. (2011). Behavioral profiles of affected and unaffected siblings of children with autism: Contribution of measures of mother–infant interaction and nonverbal communication. *Journal of autism and developmental disorders*, *41*(3), 287-301.

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2*(5), 255-267.

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, *39*, 34-50.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Sahyoun, C. P., Belliveau, J. W., Soulières, I., Schwartz, S., & Mody, M. (2010). Neuroimaging of the functional and structural networks underlying visuospatial vs. linguistic reasoning in high-functioning autism. *Neuropsychologia*, *48*(1), 86-95.

Sahyoun, C. P., Soulieres, I., Belliveau, J. W., Mottron, L., & Mody, M. (2009). Cognitive differences in pictorial reasoning between high-functioning autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *39*(7), 1014-1023.

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *Jama*, *311*(17), 1770-1777.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of autism and developmental disorders*, *10*(1), 91-103.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., ... & Leotta, A. (2007). Strong association of de novo copy number mutations with autism. *Science*, *316*(5823), 445-449.

Shephard, E., Milosavljevic, B., Pasco, G., Jones, E. J., Gliga, T., Happé, F., .Johnson, M. H., & Charman, T. (2016). Mid-childhood outcomes of infant siblings at familial high-risk of autism spectrum disorder. *Autism Research*.

Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder.*American Journal on Mental Retardation*, *105*(4), 269-285.

Stone, W.L., Hoffman, E.L., Lewis, S.E., Ousley, O.Y., 1994. Early recognition of autism. Parental reports vs clinical observation. Arch. Pediatr. Adolesc. Med. 148, 174–179.

Szatmari, P, et al. "Prospective Longitudinal Studies of Infant Siblings of Children With Autism: Lessons Learned and Future Directions." *Journal of the American Academy of Child & Adolescent Psychiatry* 55.3 (2016): 179-187.

Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X. Q., & Feuk, L. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature genetics*, *39*(3), 319-328.

Thomas, M. S. C. (2003). Multiple causality in developmental disorders: Methodological implications from computational modelling. *Developmental Science*, 6 (5), 537-556.

Thomas, M. S., Baughman, F. D., Karaminis, T., & Addyman, C. (2012). . Modelling developmental disorders.

Thomas, M. S., Davis, R., Karmiloff-Smith, A., Knowland, V. C., & Charman, T. (2015). The over-pruning hypothesis of autism. *Developmental science*.

Thomas, M. S. C. & Johnson, M. H. (2006). The computational modelling of sensitive periods. *Developmental Psychobiology*, 48(4), 337-344.

Thomas, M. S., & Knowland, V. C. (2014). Modeling mechanisms of persisting and resolving delay in language development. *Journal of Speech, Language, and Hearing Research*, *57*(2), 467-483

Thomas, M. S., Knowland, V. C., & Karmiloff-Smith, A. (2011). Mechanisms of developmental regression in autism and the broader phenotype: a neural network modeling approach. *Psychological review*, *118*(4), 637.

Thomas, M. S. C., Ronald, A., & Forrester, N. A. (2011). Modelling associations between levels of description: what can gene-behaviour associations tell us about cognitive process. *Manuscript submitted for publication*.

Tunc, B., Ghanbari, Y., Smith, A. R., Pandey, J., Browne, A., Schultz, R. T., & Verma, R. (2014). PUNCH: Population Characterization of Heterogeneity.*NeuroImage*, *98*, 50-60.

Tyszka, J. M., Kennedy, D. P., Paul, L. K., & Adolphs, R. (2014). Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cerebral cortex*, *24*(7), 1894-1905.

Uddin, L. Q., Supekar, K., & Menon, V. (2013). Reconceptualizing functional brain connectivity in autism from a developmental perspective.

van der Meer, J. M., Oerlemans, A. M., van Steijn, D. J., Lappenschaar, M. G., de Sonneville, L. M., Buitelaar, J. K., & Rommelse, N. N. (2012). Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(11), 1160-1172.

Van Lang, N. D., Bouma, A., Sytema, S., Kraijer, D. W., & Minderaa, R. B. (2006). A comparison of central coherence skills between adolescents with an intellectual disability with and without comorbid autism spectrum disorder.*Research in developmental disabilities*, *27*(2), 217-226.

Virkud, Y. V., Todd, R. D., Abbacchi, A. M., Zhang, Y., & Constantino, J. N. (2009). Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *150*(3), 328-334.

Vissers, M. E., Cohen, M. X., & Geurts, H. M. (2012). Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neuroscience & Biobehavioral Reviews*, *36*(1), 604-625.

Volkmar, F.R., Stier, D.M., Cohen, D.J., 1985. Age of recognition of pervasive developmental disorder. Am. J. Psychiatry 142, 1450–1452. Werner, E., Dawson, G., Osterling, J., Dinno, N., 2000. Brief report—Recognition of autism spectrum disorder before one year of age: a retrospective study based on home videotapes. J. Autism Dev. Disord. 30, 157–162.

Walsh, P., Elsabbagh, M., Bolton, P., & Singh, I. (2011). In search of biomarkers for autism: scientific, social and ethical challenges. *Nature Reviews Neuroscience*, *12*(10), 603-612.

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*, *127*(5), e1303-e1311.

Webb, S. J., Jones, E. J., Kelly, J., & Dawson, G. (2014). The motivation for very early Intervention for infants at high risk for autism spectrum disorders.*International journal of speech-language pathology*, *16*(1), 36-42.

Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of autism and developmental disorders*, *30*(2), 157-162.

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., & Thomas, M. S. (2007). Neuroconstructivism. *Developmental science*,*10*(1), 75-83.

Wetherby, A. M., Guthrie, W., Woods, J., Schatschneider, C., Holland, R. D., Morgan, L., & Lord, C. (2014). Parent-implemented social intervention for toddlers with autism: An RCT. *Pediatrics*, *134*(6), 1084-1093.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child development*, *80*(4), 1097-1117.

Wimpory, D. (2012) *Detection of Autism by Infant Sociability Interview (DAISI).* In: Encyclopedia of Autism Spectrum Disorders. Springer. ISBN 978-1441916976

Wolff, J. J., Gu, H., Gerig, G., Elison, J. T., Styner, M., Gouttard, S., Botteron, K. N., et al. (2012). Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. *American Journal of Psychiatry, 169,* 589-600.

Yang, J. & Thomas, M. S. C. (2015). Simulating intervention to support compensatory strategies in an artificial neural network model of atypical language development. In G. Airenti and M. Cruciani (Eds.), *Proceedings of EuroAsianPacific Joint Conference on Cognitive Science*, Torino, Italy, September 25-27th, 2015

Young, G.S., Merin, N., Rogers, S.J., Ozonoff, S., 2009. Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. Dev. Sci. 12, 798–814.

Yirmiya, N., & Charman, T. (2010). The prodrome of autism: early behavioral and biological signs, regression, peri-and post-natal development and genetics. *Journal of Child Psychology and Psychiatry*, *51*(4), 432-458.

Zachor, D. A., Ben-Itzchak, E., Rabinovich, A. L., & Lahat, E. (2007). Change in autism core symptoms with intervention. *Research in autism spectrum disorders*, *1*(4), 304-317.

Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., & Kasari, C. (2015). Early screening of autism spectrum disorder: recommendations for practice and research. *Pediatrics*, *136*(Supplement 1), S41-S59.

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural brain research*, *251*, 133-146.

Zwaigenbaum, L., Bryson, S., Lord, C., Rogers, S., Carter, A., Carver, L., ... & Fein, D. (2009). Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants. *Pediatrics*, *123*(5), 1383-1391.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International journal of developmental neuroscience*, *23*(2), 143-152.

Zwaigenbaum, L., Thurm, A., Stone, W., Baranek, G., Bryson, S., Iverson, J., ... & Rogers, S. (2007). Studying the emergence of autism spectrum disorders in high-risk infants: methodological and practical issues. *Journal of autism and developmental disorders*, *37*(3), 466-480.